

Interdependence means hardship

Global gloom about the financial outlook, not misplaced, will be relieved only slowly and by a collective will to make interdependence work.

WHATEVER happened to the optimism with which this decade began, a couple of hundred days ago? It is not just the possibility that the stand-off in the Gulf will be resolved only by open warfare, and that the Congress of Europe arranged for Paris in November will not yield the promised framework of European security, but that the world has run out of money. And the money problem is potentially the most serious because it is the most enduring. Recently acquired good sense may stave off conflicts in Europe, perhaps even in the Middle East, but the shortage of money will touch everybody everywhere. Hopes that science budgets can grow, for example, may have to be postponed. Even the new French budget is not so generous (see page 318) as to be immune from a sharp bout of inflation.

The seriousness of the problem is clear from last week's meeting of the International Monetary Fund in Washington. As oil prices have increased, poor countries have become less able than ever to pay their debts, while even rich countries are running into trouble. (People will pay more for the same goods, which is inflationary, but at the expense of their purchases of other things, which is deflationary). Japanese companies have been repatriating foreign assets on a huge scale to cover their recent losses on the Tokyo Stock Exchange (down 40 per cent this year). The United States will not even dent its chronic budget deficit before the elections in November, but meanwhile is having to spend money like water in the Gulf. And the major European states, torn between infrastructure repair (and getting their homeless off the streets) and making up for 70 years of misinvestment to the east, are prevented from following either course by high interest rates and the prospect of recession, already taking hold in Britain. What has gone wrong?

Burden

A mere shortage of money is not the explanation. Money is manufactured on printing-presses and die-stamping machines, the raw materials for which are plentiful. But there is a dire shortage of physical resources, for which money stands proxy. This is most obvious in Eastern Europe and the Soviet Union, where most industrial plant, built by the efforts of oppressed communities over four decades, is so uncompetitive with that of the rest of Europe, of which those communities now see themselves a part, that it will have to be replaced.

At \$100,000 a single workplace, about the going rate in the West, it will take a lot of investment to keep 200 million people at work. Germany seems prepared to shoulder the burden of what was East Germany, but the bills have not yet come in. Who will help out with the rest of deserving Eastern Europe? The European Development Bank, which makes a start in January, will help, but only to the extent that its members can lend it physical resources.

The West's problems are not all that different. The US budget deficit (underestimated at \$150,000 million) is a measure of excess current consumption, hitherto financed indirectly by suppliers (but the Japanese will not now be able to be so obliging as to buy a Rockefeller Center every now and again). The cost of rescuing the secondary banking system (called the savings and loan scandal) will be greater, but will temporarily be met by the printing presses (manufacturing bonds, not dollar bills), spreading claims on real resources over the future. Yet in the United States, as elsewhere in the West, the bankruptcy of industrial companies is one of the few economic activities now booming. Each industrial plant that goes bust represents an unrequited investment of resources. The waste may be the price of making the market system function efficiently, but there is now so much of it that the price has become uncomfortably high. And while the chief losers are usually individuals, the banking system's provisions for bad debts in the past year show that those with stakes in pension funds and insurance policies will eventually share the misery.

Waste

Western Europe is not blameless, but culpable in different ways (except that the British inflation rate, now 10.7 per cent a year, means the same as the US budget deficit: excess consumption). The peculiarly European way of wasting physical resources is the protection of local industries, agriculture notoriously, but an increasing range of high-technology products as well. What sense can there be in investing European resources in new enterprises with the objective of producing locally goods and services in whose provision others have already invested, and which can be bought at lower prices than they can be locally produced? This applies as much to New Zealand lamb as to computer printers from Japan. Yet even virtuous Germany seems content to go along with such



recipes for waste. Britain, often Europe's black sheep, is the most conspicuous exception in the European Communities.

Why have all these troubles surfaced more or less at the same time in different places? There is no single cause, but there is a common reason why local difficulties have grown almost beyond control. For the past two decades, the West has been congratulating itself that the growth of world trade (at an aggregate annual rate of more than 10 per cent) has made most of the world interdependent. It has been a heartening message, to which people have listened closely, looking for benefits of two kinds — a more efficient division of labour on Adam Smith's principles, and so greater personal prosperity, and a way of smoothing out temporary fluctuations of economic performance here and there. Sadly, the message has also been taken as a promise that the present state of affairs can only get better, and steadily. What now threatens is that things may get worse, and all at once.

Why should that be? People, companies and even governments have persisted in behaving as if interdependence were an abstract notion, not the reality that it is. Belief in the global market has thus given every start-up computer manufacturer the conviction that it has IBM's strength in its grasp, usually with disastrous consequences. Governments have embarked on radical domestic policies without calculating the consequences elsewhere. Chauvinism persists, built around indices such as Gross Domestic Product per capita, even though the benefits of global interdependence are that individual enterprises (and those who work for them) may best succeed even when based where overall performance is only modest. Schemes for forgiving poor countries' debts proliferate (Mr John Major, the British Chancellor of the Exchequer, invented another only the other day), but the poor countries themselves would prefer the freedom to sell goods in rich markets now closed to them. Yet the available resources are everywhere ill-matched to people's expectations.

Compromise

What is to be done? The best hope is that the way that financial institutions around the world are going down like dominoes will frighten governments into a successful negotiation on a new trading regime (called the Uruguay Round of the General Agreement on Tariffs and Trade, or GATT). There are signs of a compromise between the European Communities and the United States on agriculture, but it is insufficient. In Europe, the goal should be an informal extension of community rules and benefits to the East, putting formal membership on the back burner. Opening rich markets to poor countries is another must, however painful for the still rich. But will not steps like these impoverish the well-to-do? Of course, in the short run. But that is preferable to the impoverishment that will otherwise be accomplished by the combination of credit squeeze, inflation and the devaluation of assets now under way. □

Pugwash looks ahead

Pugwash has survived the cold war and its ending, but there is a lot remaining for it to do.

DOES an organization dedicated to the resolution of the cold war based on 1950s nuclear weapons technology have a role now that the cold war has ended? That was the most obvious question facing the Pugwash organization at its annual conference last week near London. The answer is taken to be self-evidently negative in the Pugwash Council's customary post-conference statement. Pugwash points out that negotiations on a treaty on strategic arms "have dragged on for five years" and that a comprehensive test-ban treaty is not now on the agenda — although the fuss at the recent Non-Proliferation Treaty conference (see *Nature* 347, 213; 1990) may promote it there by January. Then, the council says, with the bilateral superpower treaty as a starting point, there is now a chance to aim at the elimination of all chemical weapons. And why not now a world without nuclear weapons also?

Pugwash believes it has a future, and there is no room for dissent. Indeed, Pugwash might have added that the impending negotiation of conventions on the greenhouse gases, which raise problems similar to those in arms control, could use its members' skills. But Pugwash has outlived some of the excitements of its early days. Then, its annual meetings were the chief means by which knowledgeable people from the two opposing sides in the cold war could trade opinions. (To their credit, there have always been enough courageous people on both sides to keep the conversations alive.) Now, both military scientists and retired generals roam freely about, swapping tales of how the cold war seemed to them. (Mr Robert McNamara, US Secretary of Defense during the Kennedy presidency, enthralled last week's meeting with an account of a seminar held earlier this year in Moscow to learn what could be learned from the Cuban missile crisis of 1962.) So has Pugwash lost its function as a Trojan horse?

The organization's most valuable technical work is now done in small workshops and seminars separate from the annual meetings. They are invaluable still, and deserve the security of assured sources of finance, which would not in present circumstances be compromising either to donors or recipients. But there is a second and more difficult route that Pugwash should be exploring. If the potential foci of future trouble have shifted to places such as Iraq, should Pugwash be looking for ways of making bridges to people there inclined to take a more sober view? Or of dragging Israel, India and Pakistan into discussions of the wisdom of seeming nuclear powers? These tasks may be even more difficult than what Pugwash has done already, but they are none the less important on that account. □

■ This week's Commentary provides another view on "Strategic defence after the cold war". Page 326.

New evidence emerges in Tufts misconduct case

■ Grant application yields check

■ Secret Service called in again

Washington

INVESTIGATORS at the National Institutes of Health (NIH) have found evidence of data fabrication and falsification in the misconduct case of Tufts University immunologist Thereza Imanishi-Kari, according to sources briefed by the NIH Office of Scientific Integrity (OSI).

Although a final report on the case is still expected to be weeks if not months away, the news is the first hint of an outcome in what has become one of the longest-lasting and most divisive misconduct cases in US science. In more than three years of investigation and review, the case has drawn unprecedented attention and debate, in large part because the involvement of Nobel laureate David Baltimore, president of Rockefeller University.

The disputed work was done with Baltimore's collaboration, when he was at the Massachusetts Institute of Technology (MIT), and he was a co-author of the 1987 *Cell* paper in which the data appeared.

According to the sources, OSI investigators, using statistical analysis, forensic evidence and consultation with other scientists, have determined that all three tables in the *Cell* paper contain fabricated or falsified data. OSI has found that six of the paper's seven figures are also in some way suspect, sources say. OSI officials, through a spokesman, declined to be interviewed.

Following a still-unscheduled final review by Imanishi-Kari and her lawyer Bruce Singal, OSI will finish its report on the case. The draft will be reviewed by the NIH director and other principals in the case, then released. Should the report's conclusions be in line with the latest evidence, the US attorney's office in Baltimore, Maryland, which has been conducting a parallel investigation, is expected to press for an criminal indictment.

In that case, charges would be expected to fall within the legal definition of US Code 1001, which includes lying to a federal agency. Other charges could include mail fraud (if fabricated data was mailed to NIH), obstruction of justice and perjury. The last charge would be based on Imanishi-Kari's congressional testimony concerning the challenged paper. In four hearings, the last held this May, Representative John Dingell (Democrat, Michigan), chairman of the oversight and investigations subcommittee of the House of Representatives Energy and Commerce

Committee, has attempted to get to the bottom of the case.

Subcommittee staff conducted their own investigation, enlisting the Secret Service to analyse forensically data tapes and notebooks from Imanishi-Kari's laboratory. They concluded this year that much of the data was "not authentic", and that deliberate fabrication appeared to be in evidence.

Since Dingell's last hearing, OSI has commissioned the Secret Service to carry out further forensic analysis of the data tapes. That work is said to have gone beyond the analysis commissioned by the Dingell staff, and has found that not only were the tapes not created on their purported dates, but that some were actually produced before the experiments described in the notebooks were conducted.

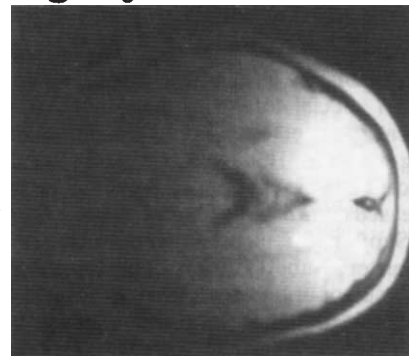
Another unexpected development in the investigation was the discovery of a grant application by Imanishi-Kari containing original data from some of the cell lines described in the disputed notebooks. Sources say the independent data provide a piece of the jigsaw puzzle that confirms evidence supplied in 1987 by whistle-blower Margot O'Toole, who had served as a postdoc in Imanishi-Kari's laboratory.

The grant application of 1 February 1985 includes a list of 150 antibody-producing hybridomas. Each was given a unique five-character code name, some 130 of which also correspond to cell lines in notebook data supplied to Congress and NIH by Imanishi-Kari as evidence supporting her paper. Because the application postdates the experiments of questioned authenticity and was found in NIH files, rather than being provided by Imanishi-Kari after the dispute occurred, the data have supplied the first independent check of the experimental techniques used in the preparation of the *Cell* paper and are said to have verified key elements of O'Toole's assertions.

Once OSI issues its report on Imanishi-Kari, it is expected to turn to the second and even more controversial aspect of the case — the 'who-knew-what-when' investigation, according to sources.

Imanishi-Kari says that she and her lawyer have been trying to schedule a meeting with OSI for months, but they have been repeatedly delayed. "My impression is that they are not too eager to see me." In the meantime, she says, "the worst of this is the agony of waiting for

High- T_c brain scan



THIS is the first image of a human brain produced by a low-field magnetic resonance imaging (MRI) scanner using a high-temperature superconductor pickup coil, rather than the conventional copper. The yttrium-barium-copper oxide coils were developed by ICI Advanced Materials, and fitted to a scanner at Hammersmith Hospital, London, by researchers from GEC.

In the scanner's 0.15-tesla magnetic field, and at liquid-nitrogen temperature (77 K), the superconductor coil has significantly greater sensitivity than copper. This gives a higher signal-to-noise ratio, which may decrease the need for repeated scans and so reduce costs. Alasdair Hall from GEC, whose brain is pictured, says superconducting ceramic coils were the logical next step after experiments to improve the sensitivity of copper coils by cooling in liquid nitrogen.

The present generation of high-temperature superconducting ceramics is so far limited to applications where their low current carrying capacity is not a problem, such as the MRI coils and sensitive aerials for military communications equipment.

P.A.

■ A 'record' T_c claim, page 321.

them to decide. When I look at this in retrospect, maybe it would have been better if it had gone straight to [the] Justice [Department]. There you have due process and rights."

Although many of the allegation of research misconduct in Imanishi-Kari's MIT laboratory arose in 1986 and 1987, internal investigations and reviews at MIT and later at Tufts consistently found no reason to retract the paper, acknowledge fabrication, or even admit gross error, with the exception of two short corrections providing some additional data (one of which is now also suspect) later published in *Cell*.

"If OSI reaches the conclusion that there was misconduct on my part, then you have to conclude that MIT covered up and Tufts covered up," says Imanishi-Kari.

Results from the second part of the investigation are not expected before the end of the year.

Christopher Anderson

Keeping faith in technology

Paris

TECHNOLOGY and industrial research are given special attention once again in a French budget that otherwise contains few big surprises. But spending in all sectors, except atomic energy research, will be increased comfortably above inflation.

Support for technology, says research minister, Hubert Curien, "puts the accent on a sector upon which the nation's future depends in the face of international competition" in a "difficult international and economic environment".

The overall civil research budget is up by 7.3 per cent from last year to FF48,672 million (\$8,691 million). If new incentives offered to industry are taken up, such as tax relief, venture capital and grants for "strategic technology" development, the research minister expects spending on research and development to reach 2.45 per cent of gross domestic product (GDP) in 1991, against 2.38 per cent last year.

The Mitterrand administration has set itself a target of 3 per cent of GDP for research spending, although the deadline has become vague. Government figures show France in front of Britain, but still behind West Germany by this index.

Basic research organizations themselves decide how to divide up their 1991 allocations. The largest, CNRS (the national centre for scientific research), will receive a 7.1 per cent increase to FF11,063 million. The medical and health research agency (INSERM) is to get a 10 per cent increase, to FF2,010 million, but will also administer an extra FF110 million earmarked for AIDS research.

But the government has placed some restrictions on the spending options of the national agencies. CNRS is asked to give priority to cognitive science, biological macromolecule engineering and the environment. The Pasteur Institute will be opening a new building for retrovirus research, focusing on AIDS.

The budget is also designed to inject young blood into an ageing population of researchers. Until recently, doctoral training was long, and competition for research posts intense. Now, doctoral research grants are to be awarded for two years and renewable for a third. "PhD theses must be short", says Curien.

Next year, 319 researcher and 330 engineer, technician and administrator posts will be created, a growth of about 4 per cent. Also, a set of measures is to be introduced to shuffle researchers upwards, giving a less uneven spread of ages across grades. The age limit for entry to the lowest research grade (CR2) is to be set at 36, falling to 31 years by 1995. All researchers over 40 in the CR2 grade will be promoted to CR1 next year. The ratio of senior to junior grades will be adjusted

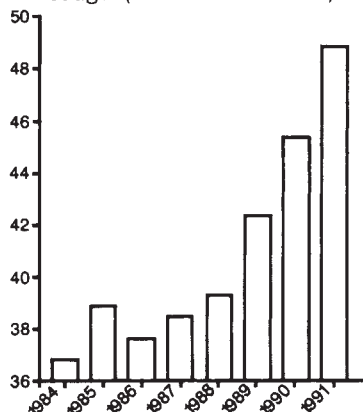
to increase numbers at the top.

But the accent in this budget is firmly on technology and industrial research. National measures include raising the threshold for tax relief on research investment to FF40 million per company (previously FF5 million and FF10 million for small and large companies respectively). A partnership scheme for doctoral training shared between industry and the state will also be expanded, while ANVAR, the national agency promoting technology innovation, gets a FF150 million boost in 1991.

France's leading role in Europe's high-technology effort is also evident. France is involved in 156 of the 386 EUREKA programmes, with the government contributing about FF750 million to the FF43,000 million national bill. Curien regards high-definition television as an essential element in Europe's effort to keep its electronics industry alive.

Another favoured project, JESSI (Joint European Submicron Silicon Initiative), could get extra French money next year.

France's civil research budget (FF thousand million)



Now that Philips has pulled out of JESSI's static memory programme, it looks likely that the Franco-Italian electronics giant, SGS-Thomson, will step in.

The national space research centre (CNES) will get an extra 13 per cent next year (bringing the total budget to FF8,119 million). Meanwhile, France has new commitments to the European Space Agency in 1991: 19 per cent of the data relay transmission module (DTRM) and 22.3 per cent of the second European remote-sensing satellite (ERS 2). The space plane Hermes, the heavy-lift launcher, Ariane V, and the Columbus contribution to the US space station Freedom come to the end of their study and definition phase in 1991. If governments agree to go ahead with construction, as expected, France will similarly have a lion's share. Next year, Hermes will cost France FF923 million and Ariane V, FF1954 million.

Peter Coles

Down from the mountain

New Delhi

FOUR chunks of rock brought last week from the Himalayas may be a step towards resolving the "case of the peripatetic fossils", which has pitted the Australian geologist John A. Talent, of Macquarie University against Vishwa Jit Gupta, professor of geology at the Panjab University in India. The question at stake is whether fossils that Gupta claims were collected in remote regions of the Himalayas are genuine, or instead came from museums and curio shops. According to Talent, the result of 20 years of publication by Gupta is that "the palaeontological literature on the Himalayas has become shot through with disinformation" (see *Nature* 338, 613; 1989; 343, 305; 343, 405; 1990).

The new specimens will provide key evidence in the scientific investigation launched by the Panjab University in Chandigarh into the alleged fossil fraud. "We will very soon be able to establish the truth and resolve the controversy", said Dr A. S. Paintal, leader of the team that has returned from the one-week expedition to the Himalayas. The expedition was proposed by the Indian National Science Academy, organized by the Panjab University and financed by the University Grants Commission.

Paintal, a fellow of the Royal Society, is a physiologist and head of the Indian Council of Medical Research. He was chosen to lead the team in his capacity as president of the Society for Scientific Values, an independent organization created in order to check misconduct in scientific research. Two of the team members are from the Geological Survey of India and four are geologists from Panjab University. Gupta, who was unable to join the team because of poor health, was represented by N. Kocher, his former research collaborator, whose main function was to guide the team to the areas where Gupta claimed to have collected the fossils.

Paintal said the aim was to collect as many specimens as possible from disputed sites. As it turned out, the team could visit only the Lahul-Spiti area in the state of Himachal Pradesh at an altitude of about 40,000 feet. Whether the rocks have embedded fossils that can be compared with those in Gupta's possession will be known only after laboratory tests begin.

Meanwhile, the team is planning to go ahead with chemical analysis and dating of the rocks as well as of Gupta's own fossils. Paintal, who expects that foreign scientists will participate in the tests on the samples, hopes the investigation will be complete by October.

K. S. Jayaraman

HIV INFECTION

Haemophiliacs win right to sue

London & Washington

IN what could prove to be a turning point in their long campaign for compensation, more than 900 haemophiliacs infected with the AIDS virus won a legal battle last week that will strengthen their argument in court against the British government over its role in importing contaminated blood products in the 1980s.

The haemophiliacs' victory gives them access to government records that could prove essential to their case. The records, which have been kept under lock and key at the Department of Health, detail government policy on the procurement of blood products in the 1980s, and are likely to shed light on at least one of the chief allegations; that despite knowing about the risk of HIV contamination, the government continued to import clotting factor from the United States rather than boosting domestic supplies.

The haemophiliacs also allege that the government was slow to introduce heat treatment for domestically produced clotting factor and screen high-risk domestic donors after 1985. By then HIV had been established as the cause of AIDS and a blood test was available.

Throughout the compensation campaign, the Department of Health has maintained that its blood product records are protected from disclosure by public interest immunity, a position upheld last year by the High Court.

After the hearing, the United Kingdom's Haemophilia Society urged a rapid out-of-court settlement. But with the Secretary of State for Health, Mr Kenneth Clarke, opposed to such a move, a full trial is likely. Clarke thinks that the documents will not help the haemophiliac's case.

By 1985 about 1,200 haemophiliacs in Britain had become infected with HIV, largely from contaminated clotting factors imported from the United States. Of these, 962 are seeking compensation, 210 have developed AIDS and 140 have died. So far the British government has made *ex gratia* payments of £20,000.

Gaining access to records important to compensation claims has also proved difficult for haemophiliacs in the United States where the targets of legal action are pharmaceutical companies. Unlike government departments, the companies, which include Alpha Therapeutics, Highland Therapeutics and Armour Pharmaceuticals — all subsidiaries of international corporations — are not bound by the Freedom of Information Act.

So far, the handful of US haemophiliacs who have filed negligence suits against some of these companies have met with little success. In one notable case, a claimant in Georgia succeeded in convincing a jury, only for the jury's verdict to be

reversed by a judge. A common stumbling block for US claimants has been proving that they can trace the source of their infection to any one manufacturer; every year a haemophiliac uses thousands of units of clotting factor, and these are unlikely to all come from a single company.

But a more fundamental challenge, and one facing haemophiliacs in Britain as well as the United States, is the difficulty of proving negligence when most infections occurred in the early 1980s before the cause of AIDS was known and a blood test was available.

Some haemophiliacs, and in particular those in Japan, argue that by 1983 — the year before HIV was established to be the cause of AIDS — there was already considerable evidence pointing towards non-heat-treated products as the source of infection. But Japanese health officials contend that when the haemophiliacs called for the introduction of heat treatment at that time the source of infection was far from clear. Last year about half a dozen of the 2,000 Japanese haemophiliacs thought to be infected with HIV filed negligence suits against the government and pharmaceutical companies.

David Concar & Diane Gershon

UNIVERSITY FINANCES

MPs count the pennies

London

THE House of Commons Public Accounts Committee, piqued by the near bankruptcy of University College, Cardiff, three years ago, yesterday issued a report saying that the Universities Funding Council (UFC) needs to strengthen its analysis of university financial forecasts, and to take "prompt remedial action" where necessary. The report, leaked to *The Guardian* newspaper, caused a controversy by suggesting that several British universities, including Keele and Surrey, are facing bankruptcy. This has been strongly denied by the universities concerned and by Sir Peter Swinnerton-Dyer, chief executive of the UFC.

University College, Cardiff, was rescued from impending bankruptcy by a £4.4 million government loan after persistently refusing to reduce spending in response to cuts in government funding. Its principal and other senior staff resigned. The report notes similar intransigence by other universities, but the suggestion that the UFC needs to take a stronger line was rejected by Swinnerton-Dyer, who says that the UFC's powers are limited by universities' autonomy.

One university finance officer says that the alarm over the hints of bankruptcy indicates a misunderstanding of the uni-

NATURAL HISTORY MUSEUM

"Think again" plea from Milton Keynes

London

THE controversial job-cutting corporate plan instituted by the Natural History Museum last April was unanimously condemned at a research meeting last week. Speaking at the 38th Symposium of Vertebrate Palaeontology and Comparative Anatomy in Milton Keynes on 21 September, Dr Beverly Halstead of Imperial College, London, delivered a scathing attack on Sir Walter Bodmer, chairman of the museum's board of trustees. The meeting was the first opportunity for researchers to discuss the plan collectively since details were released (see *Nature* 344, 805; 1990).

The delegates, mostly from the United Kingdom and Europe, passed a resolution calling on the Natural History Museum to rethink the plan's provisions — in particular the intention to shed 50 research jobs.

"Reduction in the range of expertise available, together with the strategy to devolve the research of the senior scientific staff away from curatorial activity, seems to strike a double blow at the rigorous taxonomic base on which all palaeontological and biological research ultimately depends".

Henry Gee

versities' financial status. Although they receive almost half of their £3,700 million yearly income through the UFC, universities are autonomous and can borrow money and incur temporary deficits.

The UFC says that a university running at a deficit for two or three years is "not a cause for concern", provided this is offset by surpluses in other years.

Swinnerton-Dyer surprised the committee in February when he admitted he could not be certain that 'another Cardiff' was not brewing elsewhere in the university system. He told the committee that 44 out of 74 institutions were predicting deficits by the end of the 1989-90 academic year, but refused to comment publicly on individual cases.

In the subsequent private meeting of the committee, Swinnerton-Dyer said that only three universities apart from the University of London were on the UFC's 'worry list'. Their names have been removed from the committee's final report, but were identified by *The Guardian* as Surrey, Keele and Cardiff. Surrey and Cardiff have since said that their finances are sound. Keele has acknowledged that it has a £2 million accumulated deficit, but says that this is planned, to finance a long-term programme of expansion.

Peter Aldhous

Two sides of the coin

Tokyo & Washington

JAPANESE plans for a \$1,000-million international project in manufacturing technology have been further delayed by the postponement, at the request of the United States, of a meeting arranged for last week.

If nothing more, the delay reflects big differences in the way research and development policy is set in Japan, the United States and the European Communities (EC). But regional chauvinism may also be involved.

The project, called Intelligent Manufacturing Systems (IMS), is for the development of the automated factories of the future, and would be supported to the tune of 60 per cent from Japan. Partnership has been invited from the United States and the EC.

The Ministry of International Trade and Industry (MITI) had planned to discuss reactions to the Japanese proposal in Tokyo last week, but the US Department of Commerce requested a further postponement of a meeting originally arranged for June. It will now take place in November, even though the EC have already put out their own proposal.

Department of Commerce officials reject Japanese complaints that they are dragging their feet. Mark Lieberman, Deputy Assistant Secretary for Technology Policy, points out that, unlike Japan, where MITI coerced some companies into the project, the United States cannot dictate industrial policy.

IMS was first proposed by Hiroshi Yoshikawa, head of Tokyo University's Faculty of Engineering, and won backing from MITI and Japanese industry in January (see *Nature* 343, 496; 1990). It will develop standardized manufacturing systems that are fully computerized from the design stage through to the retailing and distribution of the finished product. MITI proposed that the project organization should be funded 60 per cent by Japanese government and industry, with the remainder coming from Europe and the United States.

More than 60 major Japanese manufacturers have contributed funds to help launch the project. Two US companies, Rockwell International and United Technology Pratt and Whitney, also signed up, but then had to pull out after the US Department of Commerce heard that the US Society of Manufacturing Engineers was to act as the North American secretariat. The department insisted that US participation be organized instead under the terms of the US-Japan Science and Technology Agreement.

Lieberman says the original proposal would not have produced a "balanced, symmetrical, win-win situation", and

that "most of the benefits would have flowed in a direction which our US companies are not very comfortable with".

MITI officials have expressed surprise at this reaction, saying they had hoped the project would end complaints of lack of foreign access to Japanese research, particularly that carried out by government-led consortia, and provide a way to pay back Japan's "technological debt" to the West by generously providing Japanese expertise in manufacturing. But US and EC opinion does not accept that Japan is really ahead in computer-integrated manufacturing (CIM).

According to Fred Nichols of the Washington-based National Coalition for Advanced Manufacturing, US university-based research, driven by NASA and Department of Defense requirements, is more sophisticated than that of Japan, especially in visual simulation, artificial intelligence and sensor technology. These are essential talents for computerized manufacturing, together with the traditionally US skills in the large-scale software integration. And in some instances, Japanese companies are unaware of developments elsewhere.

EC officials admit that Japan clearly has a lead in the rapid conversion of design into manufactured products, but in their draft report argue that this is "due to cultural factors—stable career structures, shared objectives and experiences which promote communication, collective decision making and staff movement across departmental boundaries—which could not readily be reproduced in the West".

Rather than a central organization for IMS, the EC report proposes a decentralized project, much like ESPRIT and other EC research projects, with each research consortium financed from its region of origin (Europe, the United States or Japan). Collaboration between the EC, Japan and United States should be carried out between partners of "equal weight", the document states, and research should begin with a "pilot" project on a few "limited and uncontentious" studies once there is a detailed agreement on intellectual property rights.

Kenzo Inagaki, deputy director of MITI's industrial machinery division, claims the European proposal is now "very close" to Japan's way of thinking. But it is hard to see how MITI can merge its ideas with those of the United States and Europe. The United States is certain to oppose control over industrial policy by a central body, but just such a MITI-style body has already been put in place in Tokyo with the several million dollars contributed by industry.

David Swinbanks & Alun Anderson

Australian law finds balance

Sydney

THE Australian federal government last week drew the line at patenting human beings. But in a debate intended to amend patent law to accommodate innovations in biology, it agreed that some other forms of life should be subject to patent law.

In last week's parliamentary debate, Senator John Coulter proposed an amendment to the Patents Act of 1952 that would have prohibited the patenting of living animals, plants and microorganisms unless assessed by a committee consisting of geneticists, people specialized in ethics and representatives from consumer groups. This amendment was defeated largely on the grounds that it might prevent the patenting of genetically engineered vaccines.

Instead, an alternative amendment proposed by Senator Brian Haradine that human beings and their biological processes for generation should not be patentable inventions was accepted.

Coulter dismissed the accepted amendment as "obscure" while Haradine agreed that it "does not protect against abuse". Under this new bill, the patenting of living things will be solely at the discretion of the Patents Office staff without reference to parliament, bioethics committees, the public or the constitution.

But the Institute of Patent Attorneys of Australia, in a report for the House of Representatives Standing Committee on Industry, Science and Technology, takes an opposing stance and argues that "if the categories of patentable subject matter are altered to exclude living organisms, industry and other research organizations in Australia engaged in legitimate and worthwhile research programmes would be seriously disadvantaged".

Australia now stands somewhere between the United States and Europe in the scope of its laws governing the patenting of living organisms. In 1980, the US Supreme Court ruled that "everything under the Sun made by man" is eligible for a patent, as long as an invention is useful, novel and non-obvious. And in 1988 the world's first animal patent was awarded for an oncogene mouse developed at Harvard (see *Nature* 332, 668; 1988).

But in Europe, the European Patent Convention signed by 13 European members of the Council for Europe forbids patenting of "plant and animal varieties", which may nevertheless be registered and copyrighted. Although definition of "varieties" remains open to question, an attempt by Harvard to patent the oncogene mouse in Europe was rejected in 1989 (see *Nature* 340, 85; 1989).

Tania Ewing

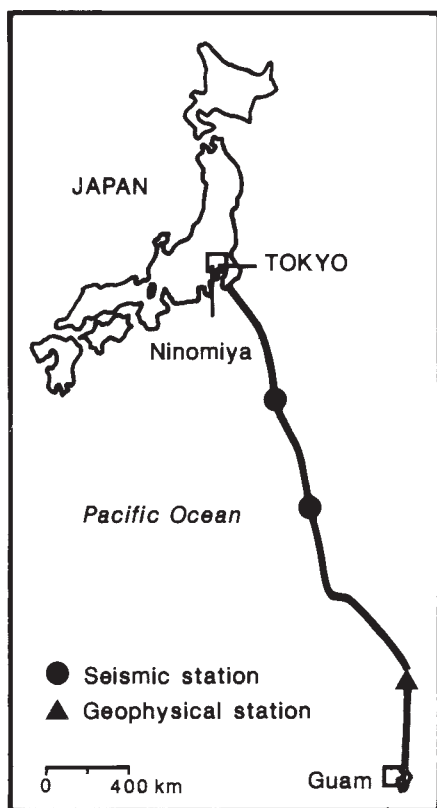
Down to the wire in the Pacific

Tokyo

JAPANESE and US seismologists have found a novel use for a discarded trans-oceanic telephone cable: they plan to make it into a seismic sensor.

Last week, the telecommunication companies American Telephone and Telegraph Co. (AT&T) and Japan's Kokusai Denshin Denwa (KDD) donated the Guam-to-Japan section of an old trans-Pacific telephone cable, TPC-1, to US and Japanese seismologists, who will use it to connect a network of highly sensitive submarine seismometers and geophysical instruments on the deep-sea floor of the western Pacific. This local set-up will eventually become part of a much larger global seismic network.

The Geophysical and Oceanographical Trans-Ocean Cable (GEO-TOC) project, first proposed in 1987, was intended to be part of the bigger POSEIDON (for Pacific Orient Seismic Digital Observation Net-



work) project, an ambitious proposal by Japanese seismologists to establish a network of 50 supersensitive seismometers on the ocean floor and on land in the western Pacific region. But POSEIDON has so far not been granted money by the miserly Ministry of Education, Science and Culture (MESC), and researchers at the Earthquake Research Institute of Tokyo University decided to go it alone with GEOTOC.

The TPC-1 cable became largely redundant in 1988 after the completion of a

new optical fibre cable (TPC-3) between the United States and Japan. But the sections of TPC-1 between Guam and Hawaii and the mainland of the United States are still in use, so only the Japan to Guam section is now available for scientific purposes. If GEO-TOC proves its worth, US and Japanese scientists hope to take over the rest of the TPC-1 cable as well as other trans-Pacific telephone cables.

The GEO-TOC network will consist of three permanent submarine stations with seismometers capable of detecting a very broad range of frequencies of vibration. One station (near Guam) will also have geoelectric and geomagnetic sensors, magnetometers, a tsunami sensor (to detect tidal waves), deep-sea current meters and other physical oceanographic equipment. Data will be relayed to both ends of the TPC-1 cable.

GEO-TOC will yield continuous real-time data from the back-arc region of the Izu-Ogasawara and Mariana subduction zones where the Pacific tectonic plate plunges beneath the Philippine plate.

The data should help to elucidate the internal structure of the Earth near the plate boundary and will also be useful for earthquake prediction. GEO-TOC, in concert with the US IRIS project and the French GEOSCOPE project, will form part of a huge global network of seismometers that will continuously monitor seismic activity around the Earth.

Japan will provide the ocean-floor instrumentation and data-processing facilities in Japan, and US seismologists hope to get funds from the National Science Foundation to establish a data-processing centre on Guam. MESC initially balked at supporting a project that involves giving foreigners free and unlimited access to 'national property', something MESC has never done before. But the Earthquake Research Institute eventually won a three-year grant of a few hundred thousand dollars this fiscal year to try out a 3-km test cable off the coast of Japan. And MESC has applied for funds in next fiscal year to begin the main project.

Junzo Kasahara of the institute estimates that the project will require about ¥2,000 million (\$15 million) over seven years — a lot of money for MESC but much less than would have been needed if scientists had laid their own cable. TPC-1 is over 26 years old and some experts fear that it may not have much life left in it. But Kasahara and other GEO-TOC supporters are confident the cable can be kept in operation for at least another ten years.

David Swinbanks

■ A new method of searching for electromagnetic precursors of earthquakes is presented on page 376 of this issue.

New highest- T_c claim

HITACHI last week claimed to have developed a new type of vanadium-based oxide superconductor with a critical temperature of 130 K. That is 5 degrees above the critical temperature of the thallium-based superconductors which have held the record for high- T_c superconductivity for over two years. The claim is of special interest because the new material, a thallium-strontium-vanadium oxide, contains no copper. So far, the only other copper-free high-temperature superconductor, barium-potassium-bismuth oxide, has a much lower critical temperature of 30 K.

But there is some scepticism about the Hitachi claim. The company says the new material has two superconducting phases with critical temperatures of 130 K and 30 K. But the Meissner effect, a key indicator of superconductivity, has been observed only in the 30-K phase. The electrical properties of the material are described as "not stable in the ambient atmosphere" in a paper submitted to the *Japanese Journal of Applied Physics*. In the past, many reports of 'unstable' superconductivity have been false alarms. D.S.

Porcine problem now

RESEARCHERS from the UK Agricultural Ministry's Central Veterinary Laboratory in Hertfordshire have transmitted bovine spongiform encephalopathy to 1 of 10 pigs injected with large doses of infected cattle brain. A second experiment to test if pigs can contract the disease by eating infected material is still in progress. The Tyrrell Scientific Committee, which advises the government on the spongiform encephalopathies, believes the announcement does not increase the risk to human health, but has recommended that the cattle offals currently banned for human consumption should be excluded from all animal feed. The ban came into force on midnight of Monday of this week, only a few hours after the announcement. P.A.

A woman's place

IN the three months since the US General Accounting Office criticized the National Institutes of Health (NIH) for favouring men in clinical trials, much has changed. First NIH reissued a policy statement, emphasizing that all deviations from a balanced gender representation in research studies must be explained. Then the Senate's reauthorization for NIH required NIH researchers to insure that their studies are statistically representative of the population at large, by both sex and race. Earlier this month Bernadine Healy was picked to be NIH's new director. And the latest development is the creation of an Office on Research on Women's Health, to be headed by Ruth Kirschstein, director of the National Institute of General Medical Sciences. C.A.

Attacking the mainline

Munich

THANKS to a US crackdown on the export of chemicals that could be used in the preparation of illegal drugs, European chemical companies may be attracting some new but unwelcome business. With prompting from West Germany, which already feels guilty over scandals surrounding the export of materials for the production of chemical weapons, officials of the European Commission are now considering concerted action to control the export of drug precursors, and chemicals which include solvents like acetone. Such action could "hand a major defeat to the drug producers", according to Gene Haislip, an official at the US Drug Enforcement Administration (DEA). Illegal drug producers require regular shipments of large quantities of the substances, all of which are so-called 'dual-use' chemicals required for the production of legal substances as well as of cocaine, amphetamines and other illegal drugs (see table). West Germany will be the first European country to take action, with a strict export law to restrict access to 12 substances expected to be introduced into parliament during next year. The law would require that export licences be granted by the government before any of these chemicals can be shipped.

Until now, it has been impossible to prosecute brokers who have bought the substances from unwitting producers and sold them to illegal drug producers in Colombia. The brokers "definitely" include German nationals, among others, says Haislip. But other chemical producers in Europe oppose introduction in the European Communities (EC) of the West German licensing system because they feel it would be ineffective. "It is just too easy to circumvent the licensing requirement by falsifying documents", says an official for the EC customs directorate in Brussels, "we only have the staff to check

one per cent of the exports". The opponents of licensing, who are thought to include France, Britain and the Netherlands, would prefer a system of registration of exports with the authorities, who would then investigate suspicious cases. But Haislip thinks this system would not be as efficient as a licensing system, which has already been shown to work with other drugs. Haislip says that the United States and Europe, working together, managed to stop the export of the chemical methaqualone, a precursor for the tranquillizers known as quaaludes. The export of quaaludes from Colombia to the US then fell by 99 per cent. "We cut off their water and they never could find the faucet", boasts Haislip.

Haislip also favours careful scrutiny of the buyers of the chemicals. DEA cut off US shipments of the chemicals to a remarkable 70 per cent of buyers in Colombia after determining that they were channelling the chemicals to illegal drug laboratories.

The proposed West German law is opposed by the chemical industry in West Germany, which does not want to be singled out for cumbersome export-licensing unless its European competitors are similarly burdened. Manfred Ritz, a spokesman for the industry association Verband der Chemischen Industrie in Frankfurt, says that the current system of voluntary controls is perfectly adequate and has even achieved some spectacular successes.

If experience with chemical weapons plant exports is any indication, the progress of the proposal into West German law, let alone European law, will be very slow. Proposed strict export laws meant to prevent a repetition of the debacle in Libya have become bogged down in the West German parliament, which will be hard pressed to pass them by the end of the year.

Steven Dickman

ILLEGAL DRUG PRECURSORS AND SOLVENTS

Compound	Illegal use	Legal use
(1) Ephedrine	Amphetamine	Bronchial dilator
(2) Ergometrine	Amphetamine	Blood pressure, migraine medication
(3) Ergotamine	Amphetamine	Blood pressure, migraine medication
(4) Lysergic acid	Amphetamine, LSD precursor	Precursor for 2, 3
(5) 1-Phenyl-2-propanone	Amphetamine precursor	Amphetamine precursor
(6) Pseudoephedrine	Amphetamine precursor	Bronchial dilator
(7) Acetic anhydride	Heroin synthesis	Common solvent; sulphonamide, vitamin B6 and aspirin synthesis
(8) Acetone	Cocaine synthesis	Common solvent; indigo and perfume synthesis
(9) Anthranilic acid	Designer drug synthesis	Indigo and perfume synthesis
(10) Ethyl ether	Cocaine synthesis	Acetylating agent, dyestuff synthesis
(11) Phenylacetic acid	Designer drug synthesis	Penicillin, dyes and perfume synthesis
(12) Piperidine	Designer drug synthesis	Drugs, pesticides, perfume synthesis; hardens resins

Export of these 12 substances from Europe is to be restricted. The 1989 Vienna Protocol recommended stricter controls on all 12 substances and their salts.

Going by the Board

San Francisco

DESPITE significant faculty opposition, the University of California (UC) Board of Regents voted on 21 September to renew for another five-year term the university's contract to manage the Los Alamos and Lawrence Livermore National Laboratories, home to much US nuclear weapons research.

UC receives about \$12 million a year from the contracts with the Department of Energy (DOE) to cover the expenses of managing the Livermore and Los Alamos laboratories, as well as the Lawrence Berkeley Laboratory, which conducts non-classified scientific research.

The vote, 13 to 3 with one abstention, was to open negotiations with the DOE for a new five-year contract which will begin in September 1992.

Many members of the UC faculty opposed the decision to continue the management contract because they believe that the link to weapons development and testing is fundamentally inappropriate for a public academic institution. In a vote taken last spring on eight of the nine UC campuses, 64 per cent of the faculty felt the university should phase out its operation of Livermore and Los Alamos, while "maintaining its current cooperative relationship with the laboratories in teaching and research".

But UC president David P. Gardner last week urged the regents to continue the 47-year relationship with the laboratories, saying that it would be "particularly ironic" to relinquish the management responsibilities now, when the world situation is changing so dramatically.

Livermore is also anxious to maintain its connection with the university. Robert Borchers, chairman of Livermore's University Relations Committee, sees the negative attitude of the faculty as something "rooted in the past". He believes that the shifting focus of the laboratories' research, including a decrease in the amount of classified research being done (classified projects now constitute only 25 per cent of the Livermore's work), can only improve relations with the university and may lead to more academic collaborations.

Beyond the philosophical disagreements, some critics have argued that the university's management has been inadequate, allowing drug abuse, security and environmental problems to crop up. Responding to those concerns, Gardner claims he is developing specific recommendations for improving management procedures.

Elizabeth Schaefer

A daunting proposition

San Francisco

ON 6 November, Californian voters will be asked to decide the fate of 'Big Green', a complex environmental initiative that is to its supporters a model for an aggressive environmental policy, but which opponents say will produce uncertain benefits at enormous cost.

Both sides last week issued reports on one of the initiative's central requirements, an enforced reduction in carbon dioxide emission of 40 per cent by 2010, and their conclusions are staggeringly different. Supporters say Big Green will save a total of \$87,000 million by 2010, while opponents claim it will cost California \$17,000 million annually.

Proposition 128, may stand or fall according to voters' perceptions of its economic impact rather than their environmental or political consciences. It specifies carbon dioxide reductions, but does not say how they should be achieved, and by painting different pictures of what Proposition 128 would demand, each side has come up with detailed projections, powerful figures and grossly different pictures of what the changes would mean to the public.

In California especially, reducing carbon dioxide emissions means persuading drivers to use their cars less, or at least to use them more efficiently. Economists at the Natural Resources Defense Council (NRDC), which prepared the analysis supporting Proposition 128, argue that by offering rebates to buyers of fuel-efficient automobiles (perhaps paid for by buyers of gas-guzzlers), by reducing the carbon concentration of gasoline and by augmenting public transport systems to decrease the miles travelled per car by 20 per cent over the next 20 years, \$9,700 million could be saved by 2000.

But Spectrum Economics of San Francisco, which prepared a report for the opposition, claims that a gasoline tax would be needed to discourage driving, and that to meet the required reductions, gas prices would rise by up to 70 cents per gallon by 2000 and twice that by 2010. Cars are not replaced often enough to change the average fuel efficiency of the California fleet except slowly, and mass-transit systems are costly, according to Steve Moss, principal author of the Spectrum report.

Electricity generation and use is another contentious issue. Chris Calwell, principal author of the NRDC report, believes that by improving the energy efficiency of lighting, appliances and buildings in general, enough energy can be saved to meet the emissions reductions required by 2000. Increased use of renewable energy sources such as solar, wind and geothermal power will then meet the 2010 goal. The Spectrum report disagrees:

even after implementing conservation measures, it says, an electricity rate increase of 20 per cent would be needed by 2000, largely to fund conversion of coal-burning power plants to natural gas use.

The different economic projections reflect in part differences in the extent of the consequences the two groups consider. The NRDC report, for example, estimates savings accruing from reduced damage to air quality, ecosystems and human health at \$50,000 million over the next 20 years. Consumers, however, may find it hard to see long-term indirect benefits, but easy to fear hefty increases in electricity and gasoline prices, particularly during the present turmoil in the Middle East. Those concerns may sway voters on 6 November. Recent polls have shown opposition to Proposition 128 mounting, and it is now nearly equal to the level of support.

Proposition 128 is an enormously complex conglomerate of environmental measures that will test the voters' stamina and understanding. In addition to its carbon dioxide emissions limits, it would, among other things, phase out the sale and manufacture of products containing chlorofluorocarbons, require developers to plant a tree for every 500 square feet of new construction, provide \$200 million to protect the state's redwood forests, permanently ban new oil and gas drilling within three miles of the coast, establish a \$500-million oil-spill prevention and cleanup fund, phase out 19 pesticides known to cause cancer or reproductive harm and establish a post for an elected state Environmental Advocate to oversee and enforce environmental laws. The state's nonpartisan Legislative Analyst's Office estimates that administrative costs alone for the proposition will reach \$90 million annually.

The environmental policy bestowed on California may be decided ultimately by the comparison-shopping abilities of voters, who have several related initiatives to choose from. Agricultural interests have responded to the pesticide portion of 128 by introducing a competing initiative with less restrictive rules, and environmentalists and the timber industry have squared off with two competing forestry propositions.

As election day approaches, Proposition 128's supporters continue to hail it as an opportunity to clean up air, water and food supplies. Detractors hope to convince voters that the measure is a poorly constructed amalgam of legislation that would strangle the state's economy. "People are wary of something that contains so much", says Scott Macdonald of the No-on-128 campaign.

Elizabeth Schaefer

The French used to have a wôrd for it

Paris

IN a move to improve standards of literacy that could yet end in tears, the French are to strip away some of their beloved accents.

A major casualty of a language reform that began with a plea for spelling 'rectification' from the Prime Minister, Michel Rocard, and ended with recommendations* ratified by the august Académie Française, is the circumflex. The 'little hat', beloved of schoolchildren and the bane of wordprocessors, will now no longer be required for the vowels 'i' and 'u'.

AOÛT, the month when Paris empties for the summer holidays, will now be a bare-headed *aout*. And scientists will be in for surprises too. (Since they are not supposed to be able to spell, perhaps it will make little difference.) We now have *croître* (to grow), *boite* (a box, as in *boite noire*), *gout* (taste), *chaîne* (as in *chaîne alimentaire* or food chain), *île*, *maitre*, *fraiche*, *voute* and *bruler*.

Some of the academy's proposals are so subtle that even the average Frenchman may neither recognize the problem nor understand the solution, but others are more radical. The hyphen is removed, whenever possible. All those Latin words beginning with *extra-*, *intra-*, *ultra-*, *infra-* and *supra-* grow shorter, making *infra-rouge*, *supraconducteur* and *ultravirus*. Ancient gallic words such as *covergirl*, *striptease*, *lockout* and *cowboy* lose their hyphens, as do *apriori*, *exvoto*, *kilowattheure*, *monoatomique*, *spatiotemporel* and *typhobacilliose*.

But for amateurs of the double-barrelled, all is not lost. *Aide-mémoire*, *serre-livre* (bookend), *ramasse-miette* (a device for picking up crumbs) and many others keep their umbilical link.

Accents are not being totally abolished. A few have even been added, especially to foreign words — a lexical equivalent of immigration papers: *artéfact*, *duodénium*, *placébo*, *édelweiss*, *pénalty* and *révolver*, for example. Finally, "diverse anomalies" are mopped up by the "look and say" principle: *chariot* becomes *charriot*, *oignon* (as in soup) becomes *ognon*, *relais* (as in *rouliers*) becomes *relai* and *skunks* (skunk) becomes *sconse*. While the education minister, Lionel Jospin, hopes that the 'rectifications' will improve standards of spelling, the simplifications could also help to stop the French language being overrun by 'anglais'.

In any case, the ministry expects the new rules to percolate into use "within a year". But signs of 'la résistance' are already emerging in some circles. Peter Coles

*Les rectifications de l'orthographe. Conseil supérieur de la langue française. Hôtel Matignon. 19 June 1990.

Openness and the MRC

SIR—Some of the ambiguities of the Medical Research Council (MRC)'s peer review system have been addressed by its secretary, Dr D. A. Rees (*Nature* 347, 116; 1990), but he made no comment on its finances. In 1989, the current parliamentary grant-in-aid to the MRC was increased by as much as 53 per cent, to be spread over the next three years. Satisfaction at this generous increase was expressed by the council. However, the secretary has now stated, in a letter to staff purporting to explain unit closures, that the MRC has not received the necessary protection against inflation, nor enough provision for current pay awards.

One of the recent actions to overcome the current 'deficit' is to close two of its internationally renowned units in Cambridge. One of these, the Molecular Neurobiology Unit, set up only in 1985, last year also underwent the detailed scrutiny of the peer review process. After this lengthy and expensive procedure, five of its six groups were alpha-rated, the referees consulted found the work of "the highest international standard" and the review subcommittee found the unit to be "good value for money" and warmly recommended that the work should continue. Subsequently, the MRC Neurosciences Board ("the wider scientific perspective" noted by Rees in his letter to *Nature*) also congratulated the unit on its achievements and recommended continuation of the work in some form, even after the present director's retirement in late 1992, setting up a search committee for a successor. Amazingly, however, the council itself subsequently rejected all these positive recommendations (from the many distinguished scientists in a wide spectrum of fields serving on the board, the review subcommittee and the panel of seven referees) and voted to close down the unit in 1992. One can only speculate on what advice they received and from whence it came.

In its wisdom, the MRC's new "strategy committee" immediately decided that this unit was now a prime target for making up a deficit, unheard of until then, and have required that one-third of the total budget for salaries and expenses be axed, a process to be started immediately. One has to ask, first, why there is a current deficit, when the MRC has received an extra £82 million over three years, plus additional funds towards the pay settlement. If there is not poor financial management, a proper explanation should surely be forthcoming. When large sums of public money are invested in new ventures (as occurred here in 1985), the scientific community at least is entitled to know why work, rated highly after intensive review, is to be cut off in midstream,

without warning. Is this the way to run a business? To claim vaguely an intention to spend more on molecular neurobiology in the unspecified future (or to start up other new ventures) is hardly a convincing reason for disabling current research in this field, still underrepresented in the United Kingdom.

While members of this unit have recently been warned not to make improper criticism of council's policy or decisions, nor any that may draw the secretary into the public arena, we have to concur, Sir, with your own opinion (*Nature* 346, 684; 1990) that more openness is essential if public confidence in the MRC is to be maintained.

P. J. BARNARD

*MRC Molecular Neurobiology Unit,
University of Cambridge Medical School,
Hills Road,
Cambridge CB2 2QH, UK*

Imperial echoes

SIR—If using imperial measures were as dreadful as Ernest L. Asten (*Nature* 346, 506; 1990) asserts, there would be two inevitable consequences. First, countries that persist with this ridiculous system would languish among the also-rans of the international community. Second, citizens of those countries would clamour to go metric and seize any opportunity to do so with enthusiasm. In fact, neither prediction is borne out by experience.

Asten must be considerably perplexed that his native land which, according to his letter, hardly knows how to build a house properly, has been the economic, technological and military giant of the twentieth century. Using British measures, it put the first men on the Moon, constructed the tallest buildings on Earth and established the world's highest standard of living. Having resided in Belgium, a metric country, the United States, a non-metric country, and Great Britain, which is somewhere in between, I can reassure Asten that houses are no better built here or in Belgium than in his own country while the plumbing is a great deal better his side of the pond.

C. H. EVANS

*Blackheath Grove,
London SE3 0DH, UK*

HIV and AIDS

SIR—Peter Duesberg (*Nature* 346, 788; 1990) requested data on matched groups of homosexual men or haemophiliacs which could show that only subjects infected with human immunodeficiency virus (HIV) developed AIDS. Such data have been available for some years in several cohort studies, and are among the many reasons why clinical scientists have been

puzzled by Duesberg's stance. For example, we show below the natural history of two cohorts from our experience. Three hundred and fourteen initially asymptomatic sexually active homosexual men have been studied longitudinally since their recruitment between 1982 and 1984. They were tested for HIV antibody when testing became available in 1984, and the cohort was found to contain 144 seronegatives and 170 seropositives; information of sexual behaviour and recreational drug use had been carefully obtained at recruitment and follow-up visits, and the two groups were closely matched. Since 1984, 25 initially seronegative subjects have seroconverted: by June 1990, AIDS had developed in 62 seropositive men, and in 6 seroconverters, after seroconversion. None of the seronegative group have developed AIDS. There have been 48 deaths from all causes (see table) in the HIV-positive subjects, and none in the seronegatives.

Similarly, 41 severe haemophilia A subjects have been followed longitudinally since 1979. Testing in 1984 revealed 26 seropositive and 15 consistently HIV seronegative. Nine out of 26 seropositive subjects have now developed AIDS, but none of the seronegatives. Mortality in the two groups from all causes is 7 out of 26 for the HIV-positive, but none in the HIV-negative group. These data clearly support a pathogenic role for HIV infection.

Outcome	HIV positive	HIV negative
Homosexual men	194	120
AIDS	68 (35%)	0
Deaths*	48 (25%)	0
Haemophiliacs	26	15
AIDS	10 (38%)	0
Deaths*	7 (27%)	0

* Death from all causes except suicide or accident.

ANTHONY J. PINCHING

DONALD J. JEFFRIES

J. R. WILLIAM HARRIS

*St Mary's Hospital and Medical School,
London W2 1PG, UK*

DAVID SWIRSKY

JONATHAN N. WEBER

*Royal Postgraduate Medical School,
Hammersmith Hospital,
London W12 0NN, UK*

Help wanted

SIR—I am preparing to write the first complete biography of Joseph Priestley. It will examine his religious and political life as well as his scientific endeavours. If any of your readers have or know of diaries, journals or letters by or referring to Priestley, I should be grateful if they would write to me at the address below.

RICHARD B. FISHER

*The Flat, 16 Highbury Place,
London N5 1QP, UK*

Risky arguments about chemicals

SIR—David Lindley made several important points in his News and Views article on risky arguments (*Nature* 346, 507; 1990) but still in my opinion missed the most important. Decisions about perceived risks must in the end be social or personal rather than scientific, and inevitably involve value as well as scientific judgements. People take into account in their perception of risks not only probability but also other types of uncertainty, whether the risk was imposed by others or undertaken voluntarily to achieve a desired end, whether or not the distribution of the risk is equitable, whether the event risked is catastrophic or easily borne, and the degree of public trust in any expert giving advice on the level of risk. To believe that such an assessment is any less "logical" than one that takes account only of numerical probability requires a very narrow and philosophically suspect definition of rationality¹.

Recent advances in the philosophy of science suggest that in giving expert advice, scientists can no longer claim that there is only one inference that can be argued from the available evidence, but rather that there are a number of possible inferences that can be defended by rational argument². Expert arguments are constrained not only by the evidence, but also by the overall conceptual paradigm,

methods and standards accepted within the expert community, and these constraints are culturally determined. A colleague of mine who works on uncertainty and risk claims that, historically, cranks have got it right more than often than experts. Be that as it may, the lay public has every reason to doubt whether expert advice on risks given by employees of governments, chemical companies, nuclear power authorities or electricity companies is always dispassionate. Unfortunately, the fact that an opinion is expert and expressed numerically does not guarantee its accuracy, impartiality or even probity.

As a case in point, Lindley accepts the conclusions of the Centers for Disease Control (CDC) that significant numbers of Vietnam veterans were not exposed to Agent Orange, and of the Veterans Administration that there were no unusual or excessive causes of death attributable to dioxin. It has, however, since been reported by others that the conclusions of CDC about exposure were incorrect, and that follow up studies of a particular group of Veterans, "Ranch Hands" who were definitely exposed to Agent Orange, showed statistically significant excesses of a number of cancers and birth defects^{3,4}. The Ranch Hands are the operators of the aircraft that delivered Agent Orange, not

ground troops. These conclusions are consistent with evidence of the carcinogenic and teratogenic effects of 2,4-D, 2,4,5-T and 2,3,7,8-TCDD in animal experiments, and in human occupational exposure^{3,5}. The evidence does not prove a causal link, but is strong enough to make unethical a social decision that ignores it². Social decisions about risk cannot be based solely on what experts consider probable, but on the whole spectrum of evidence, scientific and lay opinion, and the value system on which individual acceptance of risk to achieve social goals is based.

A. J. D. BELLETT

John Curtin School
of Medical Research,
PO Box 334,
Canberra 2601,
Australia

1. Toulmin, S. E. *The Uses of Argument* (Cambridge University Press, 1958).
2. Crawford-Brown, D. J. & Pearce, N. E. *Envir. Ethics* **11**, 153–157 (1989).
3. Clapp, R. W. et al. in *Agent Orange and the Vietnam Veteran* (The National Veterans Legal Services Project, 1990).
4. Zumwalt, E. R. Jr Report to the Secretary of the Department of Veterans Affairs on the Association between Adverse Health Effects and Exposure to Agent Orange (1990).
5. Steele, E. J., Bellett, A. J. D., McCullagh, P. J. & Selinger, B. *Tox. Lett.* **51**, 261–268 (1990).

No EOS rethink

SIR—G. Christopher Anderson ("Safety in numbers for Earth Observing Systems" *Nature* 346, 399; 1990) is incorrect in stating that members of an American Institute of Aeronautics and Astronautics review panel "have called on NASA officials to rethink the EOS design, arguing that the same scientific results can be achieved more reliably with the tightly-grouped 'clusters' of specialized satellites . . .".

The report of our workshop on this subject, *Mission to Planet Earth: Background and Issues*, dated 5 June 1990, does no such thing, nor do any of its earlier drafts. The specifically stated goal of the workshop, organized primarily for the benefit of congressional staff members, was to define and discuss all the issues surrounding Mission to Planet Earth. Our report, copies of which are available from the institute, neither supports nor opposes either the Earth Observing System's large platforms, clustered satellites or Earth probes; it simply identifies and discusses the pros and cons of each.

JERRY GREY

American Institute of
Aeronautics and Astronautics,
370 L'Enfant Promenade, SW,
Washington,
DC 20024-2518, USA

Rome University

SIR—We welcomed the letter (*Nature* 345, 658; 1990) from four professors at the University of Rome denouncing to the international scientific community the negative attitude of their faculty of medicine. We would like to express our support and to say that the situation they describe is not unique to Rome.

We (and many other Italians working like us as postdoctoral research fellows at US universities) have had similarly discouraging experiences while trying to find positions at Italian universities. Our research interests and achievements have rarely been appreciated. The study and research experience gained at major US universities has, in fact, turned out to be a handicap. It is humiliating to try to explain to our American advisers why it is so difficult for us to get a university position in Italy.

There are, however, Italian professors who are trying to improve the university system in Italy. We admire and support their efforts.

A. BORGIA

A. DAGA

A. MARTINUZZI

California Institute of Technology,
Pasadena,
California 91109, USA

Snow and cholera

SIR—In an otherwise excellent polemic against the views of Peter Duesberg on AIDS, Weiss and Jaffe's parthian shot repeats two commonly held though erroneous views on the work of John Snow on cholera (*Nature* 345, 659; 1990).

John Snow was indeed a genius. His achievement was to evolve an elegant, internally and externally consistent theory on the mechanisms and processes of cholera, and to offer practical suggestions for the prevention of the disease arising out of his theory. He did not, however, remove the handle of the Broad Street pump. Nor did this act stop the cholera epidemic in London¹.

The pump handle was removed by, or on behalf of, the local vestrymen and the epidemic was already waning by the time this was done. Nevertheless, his powers of persuasion, as his friend and collaborator, the Reverend Henry Whitehead, showed, prevented a recurrence. The father of the child whose excreta infected the Broad Street well between 30 August and 2 September 1854 took ill on 8 September, the day the handle was removed. He was ill and died in the same room as his daughter, and as his domestic cesspit leaked into the well it was presumably reinfected.

IAN G. JONES

Department of Community Health,
Fife Health Board,
Glenrothes KY7 5PB, UK

1. Cameron, D. & Jones, I. G. *Int. J. Epidemiol.* **12**, 393–396 (1983).

Strategic defence after the cold war

Michael Brower

The Strategic Defense Initiative has been out of the news recently, but its scientific and political proponents still exist. Should the United States be thinking differently about defence in this era of *glasnost*?

WHATEVER happened to the Strategic Defense Initiative (SDI)? Until quite recently, it was the most hotly debated military programme in the United States, a multi-billion dollar research effort that challenged fundamental tenets of nuclear strategy and arms-control policy. But in the past two years it has quietly faded from view, and today its name is rarely mentioned. SDI's financial support was reduced last year from \$4 thousand million to \$3.8 thousand million and is widely expected to be reduced again this year, perhaps even to \$3 thousand million. Even the Soviet Union no longer seems worried by the programme; last September the Soviet government removed one of the last remaining stumbling blocks to a strategic arms-reduction treaty (START) by dropping its insistence on formally linking the treaty to tight constraints on SDI research.

But supporters of SDI have not given up their ambition to deploy a strategic defence system in the 1990s. They assert that SDI has made great strides in reducing the costs and improving the technical capabilities of defence systems. And they have attempted to redefine SDI's mission to suit changing global parameters, placing less emphasis on repelling a massive Soviet attack and more on countering accidental launches or on attacks by other nations.

A case in point is the recent Commentary article¹ by Canavan and Teller, both of whom are long-time advocates of strategic defence. These authors mainly addressed the question of the cost effectiveness of space-based defensive rockets called 'brilliant pebbles', concluding that such rockets "appear to have a reasonable chance of providing significant defence in the near future". And they claimed that brilliant pebbles, although first developed to counter a Soviet threat, could provide "an even more complete defence against attacks from other quarters as well as against any accidental launch".

In my opinion, Canavan and Teller are wrong. There is little chance that brilliant pebbles could provide an effective defence against a Soviet ballistic-missile attack. The reasons are familiar: the

defence would be costly and could be defeated by relatively inexpensive Soviet countermeasures. And I believe that the likelihood of attack by other nations is so small that a defence against this threat is unjustified.

What proposals such as brilliant pebbles illustrate, more than anything else, is the aimlessness and confusion that is gripping SDI. The SDI office has always had trouble reconciling the dream of a near-perfect defence with technological realities. But the disarray has become even more apparent since the end of the cold war robbed SDI of its main reason for existence.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

All eyes down over arms — Brezhnev, Nixon (left) and Kosygin toast each other after the signing of the ABM treaty on 26 May 1972.

Under the circumstances, it is tempting simply to ignore SDI and trust it will go away, but this would not be wise. Left to its own devices, SDI could still be the wrench in the works of efforts to reduce strategic arms and cement a more peaceful US-Soviet relationship. The issue I am chiefly concerned about is the future of the anti-ballistic missile (ABM) treaty, which currently constrains US and Soviet strategic-defence activities. Several planned tests of SDI threaten to erode the treaty — a cornerstone of US-Soviet arms-control agreements — by exploiting ambiguities and loopholes in its language. Another, related issue concerns the accelerated development, particularly by the United States, of anti-satellite (ASAT) weapons, which have some characteristics in common with ABM weapons.

These issues will come to a head within the next few years if weapons now in the research stage move into full-scale development and production. Before then, the government must make critical decisions about SDI's budget, research strategy,

and testing plans based on a clear idea of where SDI stands as a research programme and, just as important, where it should be going. Why do we need strategic-defence research, and is SDI the right way to meet that need?

Almost from the start, there has been considerable confusion over SDI's research focus and strategic goals². The initial objective was clear: to create an impenetrable shield against nuclear weapons, rendering them, in president Reagan's famous words, "impotent and obsolete". This focused research on advanced directed-energy weapons (lasers and particle beams) which seemed

to offer the most promise of providing an effective defence in the long term. But as congressional impatience with the slow progress grew, SDI began tilting steadily towards technologies that could be deployed in the near term, as part of a limited, 'phase 1' Strategic Defense System (SDS). The system would rely on kinetic-energy weapons, or rocket interceptors (smart rocks) which would home in on missiles and destroy them by direct impact. Today,

more than half of SDI support is to develop technologies for early deployment.

As first outlined in 1987, the phase 1 defence concept relied principally on space-based interceptors (SBIs). Thousands of these were planned to be deployed in low-Earth orbit, where they would form the critical first line of defence, attacking Soviet missiles in their boost phase before the release of multiple warheads and decoys. The concept was hardly new: in the early 1960s the Pentagon briefly explored a boost-phase defence scheme called BAMBI, which relied on space-based rockets. In 1982 the US Air Force had dismissed a similar scheme. But by 1987, space-based interceptors had become a priority for the SDI office.

The original SBI concept was a two-stage rocket projected to weigh about 100 kilograms and carrying a 'kill vehicle' equipped with an infrared sensor for homing on a missile's bright booster flame. Five to ten interceptors would be housed in each of several thousand orbiting 'garages'. It was claimed that the interceptors

Associated Press

would incorporate propulsion and homing-vehicle technologies that were well-understood — although of course no one has ever intercepted a missile from space before. The entire system would be controlled by a small number of battle-management satellites receiving instructions from the ground.

There were several flaws in this plan. First, SBIs could be foiled by fast-burn intercontinental ballistic missiles with a boost phase of significantly shorter duration than that of current Soviet missiles. The reason is simple: the strike range of an interceptor is limited by its velocity multiplied by the time available for interception. As the boost phase is shortened, the range decreases, thus driving up the absentee ratio (the number of interceptors required in orbit for at least one to be always within range of missile launch sites). For a sufficiently short boost phase, interception becomes impossible because the rockets cannot cover the distance between their orbital altitude and the missile burnout altitude in the available time.

Second, SBI garages and battle-management satellites would be vulnerable to ASATs. As one ASAT could destroy a garage containing several interceptors — or even a battle-management satellite controlling hundreds of interceptors — the cost-exchange ratio (a measure of how easily a defence system can be overcome) would almost certainly favour the ASAT.

Brilliant pebbles

Brilliant pebbles are essentially advanced versions of the SBI, designed to lower the cost and address some of the shortcomings of the original concept. The mass of each brilliant pebble would be smaller (about 40 kilograms), thus reducing the cost of launching the defence system. The rockets would not be bunched at garages but would be deployed in 'singlets', hence raising the cost of ASAT attacks. Further, each brilliant pebble would be equipped with a powerful computer as well as sensors, enabling it to function autonomously and eliminating the need for battle-management satellites.

According to Lowell Wood of the Lawrence Livermore Laboratory, "There are literally no technical obstacles to a brilliant pebble ballistic-missile defense". The SDI office apparently agrees, for in February this year it announced it had redrawn plans for the phase 1 Strategic Defense System to rely on brilliant pebbles. More than 4,000 brilliant pebbles would be stationed in orbits of 400 to 500 kilometres. The total cost of the defence — including 1,500–2,000 ground-based interceptors — would be \$55 thousand million, about half that estimated for a phase 1 system two years earlier.

Yet the SDI office's enthusiasm for

brilliant pebbles is at best premature. In December 1989, the Defense Science Board, an official advisory group to the Pentagon, concluded the brilliant pebbles design is "by no means complete, and is still changing"⁴. And some of the claims being made for the interceptors seem to be exaggerated. Wood and others assert that each brilliant pebble would carry a computer equivalent in power to a Cray I supercomputer in a package the size of a cigarette packet. They also state that the brilliant pebble's sensors would be able to carry out all the functions of attack warning, assessment, target tracking and interception previously assigned to separate (and complex) systems. How could all this be attained at low cost using existing technologies?

More importantly, brilliant pebbles do not solve either of the two main problems with the SBI concept. Although deployed singly, they would still be vulnerable to ASAT attacks. And the ASAT — brilliant pebble cost-exchange ratio, although lower than that for the SBI, would still probably favour ASATs, for three main reasons. First, the attacker would not have to destroy all the defenders but only those that happened to be over intercontinental ballistic missile launch sites at the time hostilities began; this implies at least a 10:1 ratio (the brilliant pebbles' absentee ratio) in the attacker's favour. Second, a brilliant pebble must be placed in orbit, requiring a launch velocity of 8 to 10 kilometres per second, whereas an ASAT needs ascend only to the brilliant pebble's altitude, requiring a launch velocity of about 3.5 kilometres per second. The velocity difference implies roughly a 7:1 advantage in launch weight, and hence cost, for the ASAT. And third, brilliant pebbles must carry their propellant with them into orbit, whereas only the ASAT kill vehicle need be launched into space, giving the ASAT an additional factor of 10 or so advantage. One can argue that superior US technology as well as counter-measures such as decoys and manoeuvring could partially offset the ASAT's inherent advantages, but it is difficult to see how the advantages could be entirely overcome (see, for example, refs 5 and 6).

Further, brilliant pebbles could not handle fast-burn intercontinental ballistic missiles. Even against existing Soviet SS-25 missiles, they would be only marginally effective, according to official statements. SS-25s are solid-fuelled and burn out in approximately 180 seconds, considerably less than the 300 seconds typical of older, liquid-fuelled missiles. Studies⁵ by US missile manufacturers indicate that boost times as short as 60 seconds are feasible and would entail at most a modest (20 per cent) penalty in missile launch weight per unit payload. In the past, the Soviet Union has fielded a new generation of missiles approximately every ten years. Thus it

appears likely that fast-burn missiles could be deployed by the end of the 1990s — about the time one might expect a defence based on brilliant pebbles to become fully operational.

Obsolete

Why has the brilliant pebbles scheme been adopted wholeheartedly by the SDI office? One reason given is that it is a necessary first step to a more effective defence based on advanced technologies. But the current scheme seems destined to become obsolete almost before it is deployed, and there is no guarantee, at this stage of research, that more effective defences will be available by the time they are needed. Early deployment has been said to provide some protection for land-based strategic forces. But would space-based defence be the best way to provide this? START will do as much to reduce the Soviet intercontinental ballistic missile threat (by eliminating half of the most accurate and powerful missiles, the SS-18s) than could the phase 1 defence system, at almost no cost.

But supporters of SDI realize that Congress has little patience for a research programme that costs a lot and produces few tangible results. Consequently, they are under intense pressure to come up with a scheme that can be deployed quickly and cheaply, using 'off-the-shelf' components.

Proposals to deploy a small-scale defence system against accidental attacks or attacks from small countries take this logic to its extreme. An accidental launch protection system could consist of a single, ground-based system using fewer than a hundred rocket interceptors or of grander schemes involving hundreds of ground-based or space-based interceptors. Canavan and Teller advocated deployment of 100 brilliant pebbles at an estimated cost of less than \$1 thousand million. They claimed this would provide "a good (though never complete) defence against missiles fired accidentally or by a terrorist state". But it and most other schemes would also require the abandonment of the ABM treaty and probable Soviet withdrawal from START.

It is hard to see how the threat of accidental attacks or attacks from smaller countries could justify immediate deployment of a limited defence system. The danger of accidental or unauthorized missile launches by the Soviet Union is generally regarded as extremely low. Small countries are most likely to be interested in acquiring short- and medium-range ballistic missiles to be targeted within their regions, rather than at the United States, as recent events in the Gulf demonstrate. A space-based missile defence would be ineffective against such missiles. If terrorists, or terrorist states, were determined to unleash nuclear or chemical attacks on the

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

A Soviet missile is destroyed in accordance with US–Soviet agreements. START should lead to further reductions in strategic weapons.

United States, it would be easier for them to smuggle in weapons than to use costly and sophisticated ballistic missiles. Although Iraqi threats to use chemically armed missiles against Israel and other Middle East targets have made the world more aware of the dangers of ballistic-missile proliferation, they have not fundamentally changed this conclusion.

US strategic-defence policy is in sad disarray. Unable to fulfil the original dream of SDI, its managers and supporters have pushed for early deployment of a limited, flawed defence system in a bid to ensure the programme's survival. The government needs to re-evaluate its strategic priorities in the light of technological and political realities.

What principles should guide US policy? First, the government must recognize that US interests are best served by a strong ABM treaty. By placing strict limits on defences, the treaty has eliminated a powerful incentive to the arms race and made it possible for the United States and Soviet Union to reach agreements (SALT I, SALT 2 and START) limiting strategic arms. Although the cold war now appears to be over, neither the United States nor the Soviet Union is likely to make significant further cuts in nuclear weapons unless the treaty remains in force.

Yet the treaty's future is in doubt. Several SDI tests planned in the next few years will strain the legal limits of the treaty and in some cases may exceed them¹. An example is the flight testing of brilliant pebbles prototypes that began a few weeks ago. To avoid the treaty's ban on the testing of space-based ABM components, the prototypes will be launched from the ground, and the interceptors' homing vehicle will attempt to strike target missiles as the vehicles fall back to Earth. Although these tests may be technically legal, they will violate the spirit and purpose of the treaty by bringing the United States closer to being able to deploy a space-based defence system. In addition, since the prototypes' sensors will be able to perform the tracking and

guidance functions of a traditional ABM radar, the tests may well be a violation of the treaty's ban on the testing of mobile ABM components. In 1978, the United States and Soviet Union agreed to discuss ambiguous situations like this when they arose, but the US Defense Department has resisted doing so in this case.

Another challenge to the ABM treaty is the growing similarity of ASAT and ABM systems. The treaty permits ASATs, but because low-orbit satellites and ballistic missiles follow similar trajectories in space, there is a potential overlap between the technologies. The Soviets have one relatively primitive but operational ASAT system. In 1988, the United States cancelled its more advanced programme, the miniature homing vehicle, after Congress repeatedly barred testing of the device against targets in space. In the past year, however, the United States has redoubled its ASAT development efforts, in particular on a ground-based interceptor similar to SDI's exoatmospheric re-entry-vehicle intercept system (ERIS). More money is also being sought for various ground-based laser ASATs, many of which are direct outgrowths of SDI directed-energy research.

Even if deployed, these ASATs would not constitute an effective ABM system. Nevertheless, there is no agreed performance threshold to distinguish between the two types of system. Moreover, as both nations take steps to protect their satellites from attack, ASAT performance is likely to improve; and as nuclear arsenals are reduced, even modest ABM capabilities may gain importance. Thus, either superpower may eventually be able to test and deploy ABM systems in the guise of ASATs.

These looming challenges to the ABM treaty imply the need not only for mutual restraint in the development and testing of new technologies, but also for negotiations to clear up ambiguities and close off loopholes in the treaty. In addition, it would be in both countries' interests to negotiate an agreement prohibiting the testing and deployment of ASAT

weapons or, failing that, setting strict limits on ASAT capabilities.

Besides staying within the bounds of the ABM treaty, SDI research and development efforts should be focused on those defence concepts that are both clearly feasible and serve a valid strategic or military purpose. Ground-based systems intended to defend strategic targets such as missile silos and command posts, as well as anti-tactical ballistic missile systems that could protect US troops or allies in regional conflicts, might fit this guideline. Vulnerable and costly space-based defences intended to defend the entire country against ballistic-missile attacks clearly do not. This does not mean that there should be no research on advanced technologies that might in future defend the population. But this research — on brilliant pebbles, for example — should be kept strictly in the laboratory.

If these suggestions were followed, there would be three chief benefits. First, money would be saved. A research programme conforming to the guidelines I suggest would cost less than \$2.5 thousand million a year. Second, a strong ABM treaty would improve the chances for further serious reductions in strategic arms after START. And third, by conducting a more coherent and rational research effort, the United States would gain a better idea of the true potential of missile defence. □

Michael Brower is at the Union of Concerned Scientists, 26 Church Street, Cambridge, Massachusetts 02238, USA.

1. Canavan, G. & Teller, E. *Nature* **344**, 699–704 (1990).
2. Clausen, P. & Brower, M. *Technology Review* p. 60 (October 1987).
3. Wood, L. *Aerospace America* 20 (April 1990).
4. Defense Science Board Report on SDIO Brilliant Pebbles Space-Based Interceptor Concept (Office of the Under Secretary of Defense for Acquisition, Washington, DC, 1989).
5. Speed, R.D. ASATs vs. Brilliant Pebbles Report UCRL-ID-103669 (Lawrence Livermore National Laboratory, 1990).
6. Garwin, R.L. *Nature* **346**, 21 (1990).
7. Bloembergen, N. & Patel, C.K. *Rev. Mod. Phys.* **59**, No. 3 Part II S30–S32 (1987).
8. Bunn, M. *Foundation for the Future: The ABM Treaty and National Security* 90–103 (Arms Control Association, Washington, DC, 1990).

Clouds and global warming

Real clouds may moderate global warming by negative feedback, or may work the other way. Comparison between observations and predictions suggests that clouds are cooling influences outside the tropics.

THE question of whether clouds would substantially change the prediction of global warming caused by the accumulation of carbon dioxide in the atmosphere has been much debated, not least because there is no simple way of answering it with precision while back-of-the-envelope calculations exist on either side and are memorable, even compelling. That is why the issue has been about for twenty years.

The simplest argument is that clouds provide negative feedback for global warming. If the surface temperature of the Earth is increased, evaporation from the ocean surface will be increased, cloud formation will be enhanced and, because clouds reflect sunlight, the net receipt of energy at the surface of the Earth will be decreased. It is true, of course, that water vapour is also in its own right a greenhouse gas (which works the other way) and that, by the same simple argument, the lower surfaces of clouds will intercept long-wavelength radiation and keep its energy in the lower atmosphere, but if surface warming is the worry, the effects of the upper surfaces of clouds must be the more important.

A second envelope-back will take the argument further. Clouds are dynamic structures, are they not, and may notoriously be a way in which water vapour and water drops are carried rapidly upwards through the height of the clouds to deposit the latent heat of ice formation near the upper surface? At the lower surface, by contrast, at least if rain is not falling from the cloud, the latent heat of evaporation is balanced by that condensation. So the right kind of cloud, a thundercloud, may be an effective means of transporting energy from low to higher altitude. The negative feedback is all the stronger.

This argument can, unfortunately, be turned around. A cloud is not just a place of high water concentration lying between a lower level at which the temperature is the dew-point and an upper level (below the freezing point) determined by the amount of water in circulation, but is driven by physical movements of the atmosphere, otherwise called winds. But how to dissipate that external energy? By making the cloud into a kind of heat-pump, transferring heat from the upper to the lower surface.

The balance between these contradictory arguments would have been settled long ago if only it had been possible to

include real clouds in the models of the general circulation of the atmosphere. Sadly, that is not even now the case, and for good reason. The most elaborate climate models may treat a dozen slices of the atmosphere between the surface of the Earth and outer space as independent, and may treat the atmosphere itself as so many vertical cells with dimensions of the order of 5° or 500 km. Whoever saw clouds that big?

In reality, the modellers have been forced to account for clouds by means of a parameter called cloudiness. Atmospheric water vapour is determined by surface temperature, and is in turn the determinant of cloudiness. And the effects of cloudiness on the state of the atmosphere, especially the radiation balance within it after allowing for the role of water vapour as a greenhouse gas, is meant to be an average between the behaviour of the atmosphere when there are no clouds and that when the surface of the Earth is obscured — an average between black and white, so to speak, almost calculated to excite suspicion. How useful is that stratagem?

Luckily, one development of the past few years makes the question almost tractable — the flow of data from the satellite observing campaign known as ERBE, for Earth Radiation Budget Experiment. There are three satellites altogether — Nimbus 9, Nimbus 10 and a dedicated ERBE satellite, all in polar orbits with different altitudes. In the long run, this is how the Earth's radiation budget will be monitored; it is a sad reflection on professional people's sense of what is important that they are not clamouring to have a better system (better coverage, more frequency channels, greater precision) launched tomorrow.

Now J. T. Kiehl from NCAR (the National Atmospheric Research Center at Boulder, Colorado) and V. Ramanathan from the University of Chicago have made the best of such data as there are by attempting to compare the behaviour of a relatively sophisticated version of the NCAR "community climate model" (CCM) with the ERBE-observed radiation from the surface of the Earth (*J. geophys. Res.* **95**, 11,679; 1990). They emphasize that, at this stage, they are not seeking to estimate the importance of whatever feedback mechanism may be operating, but to compare CCM with reality. They conclude that there is little differ-

ence between model and the observations, but luckily they leave room for argument on the backs of envelopes.

Comparison between the observations and the model calculations is possible only because the models, involving as they do the transfer of energy by radiation from one level in the atmosphere to another, allow the calculation of the radiation expected by the ERBE sensors. Are the predictions very different from the observations, and if so how and where? The comparison generally rings true in the sense that the model tends to predict greater extremes of, say, outgoing long-wave radiation as a function of latitude than the observations find, but with the same general shape; that is another way of saying that the models have still to be refined. There are a few technical difficulties also; for example, it seems that some parts of the Earth are never so free from clouds that their "clear-sky" behaviour can be observationally determined.

Broadly speaking, the discovery that matters in this painstaking comparison between prediction and observation is that the uncertainties occasioned by clouds are greatest in the tropics. In most latitudes, clouds do have the effect of cooling the Earth beneath, but in the tropics there is a difference of 4 W m⁻² between the model and observation in the amount of energy accumulated near the Earth. There is nothing seriously wrong with that difference, a mere half of a per cent of incoming radiation, even if it might make a big difference over months or even weeks.

What matters more, for the time being, is that the argument suggests a few specific modifications of the climate models. A better scheme for allowing for the effect of increased surface temperature on the transport of water vapour from the oceans is an important need. The modellers seem ready, even eager, to take that on board. A rerun of the calculations should tell whether that will, for the time being, suffice.

But further down the road, the envelope-back devotees will have to be won over by a genuine and verifiable prediction. If, for example, the models say that increased surface temperature means more water vapour, what does that mean for the frequency of cloud cover? One is bound to suspect that to be a sensitive test of the accuracy with which the feedbacks can be gauged. How does it come out?

John Maddox

Schrödinger's cat ensnared

Philip Ball

JOHN BELL has called for the word 'measurement' to be banished from quantum theory¹. It could scarcely have been more ubiquitous, however, at a recent workshop* concerned with ways in which quantum-mechanical effects can manifest themselves on a macroscopic scale. For the dilemma is not how to produce such effects but how to look at them. And although the concept of measurement emerged arguably as ambiguous as ever, the discussion was clearly the richer for its inclusion.

But the most profound message from the meeting was that interpretations of quantum theory are no longer a matter of philosophical taste. Thanks to the development of electronic systems of remarkable sensitivity, we are now in a position to study experimentally the questions that Bohr and Einstein battled over for much of their lives equipped with little more than intuition².

On the whole, quantum theory can be used without having to confront such

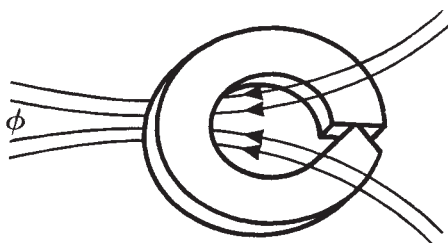


FIG. 1 Superconducting weak-link ring, containing a point contact and penetrated by flux ϕ .

issues. In the past, quantum effects have tended to occur only on the microscopic (essentially atomic) scale, and ambiguities in the orthodox 'Copenhagen' interpretation of quantum theory become unimportant when one uses classical measuring devices that probe ensembles of, say, 10^{20} or more microscopic 'quantum objects' (I. Percival, Queen Mary College, London). It is now possible, however, to create individual macroscopic quantum objects, perhaps a few centimetres in size — the same order as the classical apparatus used to probe them. Studying such systems, said Yogi Srivastava (University of Perugia), is akin to climbing inside a proton with our measuring devices.

The most promising candidates for displaying macroscopic quantum behaviour are various kinds of electronic circuits, particularly semiconductor heterostructures (such as GaAs/GaAlAs layered systems), in which the electrons behave much like a two-dimensional gas, and superconducting rings containing 'weak links',

as are used in SQUID magnetometers (Fig. 1).

Terry Clark (University of Sussex) discussed the state of the art in SQUID ring experiments. The weak link in these rings (typically fabricated from a low-temperature superconductor such as niobium) is a point contact; transport of electron 'Cooper' pairs across the contact relies on quantum-mechanical tunnelling. This is a probabilistic process, resulting in a build-up of charge on either side of the junction so that the device develops a capacitance.

At the microscopic level, charge Q and magnetic flux ϕ are related, like position and momentum, by an uncertainty principle: $\Delta\phi\Delta Q \geq h/2$. The upshot of this relation is that the weak-link ring can adopt two quantum 'modes' — a flux (inductive) mode, in which charge flows but flux lines through the ring tend to be localized, and a charge (capacitive) mode, in which charge tends to be localized. Different quantized energy states (eigenstates) of the charge and flux modes are coupled by some characteristic tunnelling frequency, so that in principle the ring may lie in a quantum superposition of states. Clark and colleagues have been probing these macroscopic eigenstates, attempting in particular to address the question of whether it is possible to catch the ring in such a quantum superposition — the equivalent of asking whether one could see Schrödinger's notorious cat³ in a superposition of 'live' and 'dead' states.

This is where the spectre of measurement creeps up through the floorboards. The Copenhagen interpretation proposes that the act of measurement 'collapses' a quantum superposition into a semiclassical wave packet (Percival). Allan Widom (Northeastern University) made the point with his claim to be able to exist quite happily in a superposition of macroscopic states — say, inside and outside the lecture theatre — so long as everyone looked the other way.

Thus, while an isolated quantum system is (in the Copenhagen interpretation) fully described by quantum mechanics, any form of coupling to a classical environment (such as a macroscopic measuring device) tends to reduce it. But the hope is that if the coupling can be made very weak, yet still sufficient to carry out a measurement, one might avoid reduction. Tim Spiller (University of Sussex) discussed possible schemes for making such measurements in idealized systems in which a well defined quantum object is coupled to a well defined classical environment. One can probe the quantum object either by measuring it directly or by

monitoring its effect on the coupled environment. The former approach is generally the more familiar, but Spiller showed that only by adopting the latter can one hope to be able to see macroscopic superpositions of quantum states.

This, therefore, is the approach used for the SQUID ring experiments. Weak coupling is achieved by using, as the classical environment, an electronic circuit containing an induction coil (with inductance L) and a capacitor (with capacitance C). When driven by an alternating current source, this so-called tank circuit has a resonant frequency given by $1/\sqrt{LC}$. But if the inductor is inductively coupled to the weak-link ring, changes in the flux state of the ring cause changes in L and thus in the resonance condition. In this way, the quantum state may be probed without looking at it directly — that is, without collapsing the wavefunction by measuring flux or charge.

The response of the tank circuit as a

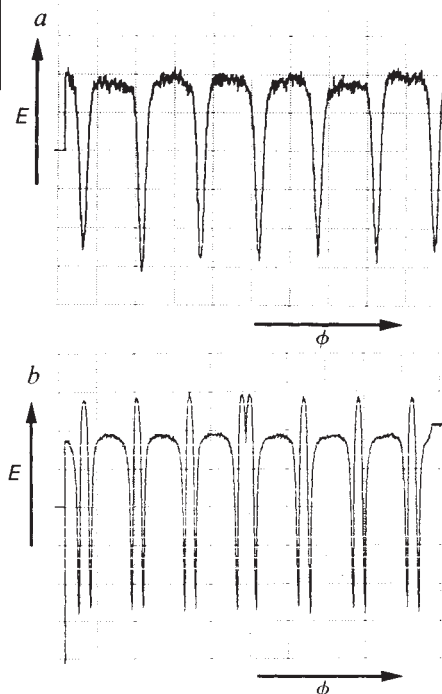


FIG. 2 a, Macroscopic ground state, and b, first excited state of the SQUID ring as a function of external applied flux ϕ . The minima correspond to eigenstates encircling an integral number of flux quanta ϕ_0 . Because of the weak link, the ring supports a screening supercurrent in response to the applied flux that fails periodically as ϕ increases, letting an additional flux quantum thread the ring. The ring passes from one flux state, $n\phi_0$, to the next, $(n+1)\phi_0$, via a quantum superposition of states. The response of the tank circuit is monitored by a radio receiver which probes the resonance condition. Remarkably, these results were obtained without alternating current driving of the tank circuit — they show the response due solely to self-excitation by thermal voltage fluctuations at a temperature of about 4 K. This ensures that coupling to the quantum object is maintained at an absolutely minimal level. Data are from T. Clark *et al.* (University of Sussex).

* Workshop on Macroscopic Quantum Phenomena, University of Sussex, 23–24 August 1990.

function of external applied flux through the ring essentially monitors the second derivative of the eigenenergies with respect to flux. The results (Fig. 2) reflect the fact that these energies are periodic in flux (when monitored in the flux mode) or charge (for the charge mode) (Clark). A theoretical analysis for the ground state shows that the minima of this periodic response correspond to macroscopic eigenstates that encircle an integral number of flux quanta $n\phi_0 = nh/2e$, whereas in between successive minima the ring is in a true quantum superposition of degenerate eigenstates. Transitions between ground and excited macroscopic states can also be monitored, and the signal from the first excited state (Fig. 2b) is consistent with expectations.

The problem of measurement is also at the heart of the Einstein-Podolsky-Rosen (EPR) paradox¹, according to which its originators claimed to have exposed the inability of quantum mechanics to provide a complete description of the world. The EPR argument may be summarized in a *gedanken* (thought) experiment in which two quantum objects (say electrons) are produced from a source in such a way that they are correlated (say, one spin has to be up and the other down). If one spin is measured after the objects have become macroscopically separated, the other is automatically determined at the same time. Quantum mechanics insists that before measurement collapsed the wavefunctions, the particles had no defined spins but were in a quantum superposition of the two states. So either one accepts the disturbing possibility of instantaneous action at an arbitrarily large distance, or one admits that quantum mechanics is incomplete and that the spins were defined all along.

Alain Aspect has devised and performed optical experiments to try to resolve this paradox², but the interpretation of these attempts is ambiguous. Claudia Tesche (IBM Yorktown Heights) reviewed the progress towards a version of Aspect's experiment using superconducting SQUIDS. Here again, the bugbear of distinguishing quantum object from environment threatens to cloud the issue; and the technological demands are daunting. The IBM group hope nevertheless to have their experiment up and running in the not-too-distant future. But in the meantime, the Bohr/Einstein controversy looks destined to run and run. □

Philip Ball is an assistant editor of *Nature*.

Why birds (still) fly south

James L. Gould

THE amazing flexibility of homing and migrating birds has been a puzzle for years. Remove cue after cue, and yet animals still retain some backup strategy for establishing flight direction. A report by Able and Able on page 378 of this issue¹ reveals a new layer of subtlety in the intertwined and redundant navigational systems of migrants.

Migration is one of the most highly programmed behaviours in birds. Even yearlings with no experience moult at the proper time, put on the fat they will need for the journey, depart at the right time of year and in the correct direction, and stop flying only when they have reached the target area. The phenomenon can be studied readily because hand-reared chicks will enter a phase of 'migratory restlessness' in a featureless laboratory at the same time as their wild-reared peers are departing on their autumn or spring journeys, and will hop in the appropriate direction night after night until about the time their migrating colleagues reach their destination.

But like many complex innate behaviours, the migration programme must be calibrated with information from the environment — day length, for instance, sets the annual clock. The navigational compass, in particular, needs careful calibration. Because most birds migrate at night (a strategy which helps prevent overheating), the cues available in flight are star patterns when the sky is clear, and the Earth's magnetic field under overcast skies. As Emlen showed 20 years ago², young birds will imprint on artificial rotating star patterns during a sensitive period, and relate this information to the axis of celestial rotation (true geographical north).

When the time for migration arrives, the birds can choose the proper direction on the basis of only an isolated, nonrotating section of the pattern. The birds clearly know the appropriate course to set with respect to true north, and use that to calibrate a star map. Overcast nights are dealt with through a magnetic compass, which appears even in birds reared in complete visual isolation³. Later, Wiltschko and Wiltschko found that birds can use their magnetic compasses to calibrate a static pattern of stars⁴, such as might be visible to young birds with only brief or partial views of the night sky during their sensitive period.

Because the north magnetic pole is located almost 1,500 km from the rotational pole, there is a substantial difference between these two compass systems in much of the world (particularly in the higher latitudes where many species pass

the summer). If the innate direction of migration is referenced to true north, calibrating a celestial compass against a magnetic one in any area with a large magnetic declination seems certain to introduce navigational errors. It seems to make more sense, when circumstances permit, to calibrate the magnetic compass against a celestial standard. Able and Bingman have found evidence of such calibration⁵ to both day and night skies, and now Able and Able report¹ their elegant experiment designed to explain how this adjustment is managed.

Able and Able reared birds through the sensitive period under artificial star patterns but in the ambient magnetic field. One group of birds saw the patterns rotating, as is normal, about a point in the north sky; the other three were exposed to an apparent celestial pole to the east, south or west. When the time for migration approached, the birds were tested individually with no celestial cues. Half of each group was left in the ambient field; among these, the pole-is-north group chose the expected north-south course, whereas the pole-is-east/west groups attempted to set off 90° to the left or right of normal. To ascertain that the birds were not using some nonmagnetic cue to guide this behaviour, the other half of each group was tested in a rotated field; in each case the migratory orientation was rotated to the same degree.

Because the Earth's magnetic field is always available, no amount of natural deprivation can leave migrants without a compass: the magnetic sense is the ultimate (though potentially inaccurate) backup system. The interesting thing about the new results is that they show that birds can have alternative (and even reciprocal) calibration systems to achieve the maximum degree of flexibility and accuracy in the face of the changing availability of cues. It is yet another reminder of the effortless interaction of instinct and programmed learning in contexts that require calibration⁶ — situations as varied as food learning, enemy recognition and, of course, navigation. □

James L. Gould is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-1003, USA.

1. Bell, J.S. *Physics World* **3**(8), 33–40 (1990).
2. Shimony, A. in *Niels Bohr: Physics and the World* (eds Feshbach, H., Matsui, T. & Oleson, A.) 285–305 (Harwood, Chur, 1988).
3. Schrödinger, E. in *Quantum Theory and Measurement* (eds Wheeler, J.A. & Zurek, W.) (Princeton University Press, 1983).
4. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **46**, 777–780 (1935).
5. Aspect, A. et al. *Phys. Rev. Lett.* **49**, 91–94 (1982).

1. Able, K.P. & Able, M.A. *Nature* **347**, 378–380 (1990).
2. Emlen, S.T. *Science* **170**, 1198–1201 (1970).
3. Wiltschko, W. & Gwinner, E. *Naturwissenschaften* **61**, 406 (1974).
4. Wiltschko, W. & Wiltschko, R. *J. comp. physiol.* **109**, 91–99 (1976).
5. Able, K.P. & Bingman, V.P. *Q. Rev. Biol.* **62**, 1–29 (1987).
6. Gould, J.L. *A. Rev. Psychol.* **37**, 163–193 (1986).

Unwinding the gordian knot

John D. S. Jones

TAKE a piece of string, tie a complicated knot in it, and then sew the two ends together so you cannot simply slip the knot off. Now, can you unknot the string? Anyone who has been fishing, with the bait on one end of the line and the reel on the other preventing you from slipping off the knot in between, will know that this can be an intricate task. And can two knots made by different processes in fact be the same even though they appear to be different? Such questions are tackled mathematically in knot theory. This theory also extends to links, which arise when you start with several pieces of string, tangle them together introducing knots and at the same time intertwining different threads, and then sew the ends together — like the horror of two fishing lines getting tangled. In his article in this issue (*Nature* 347, 367–369; 1990), H. K. Moffatt proposes to study knots and links by using techniques from fluid mechanics. This leads to a notion of the ‘energy’ of a knot or a link. Moffatt shows that this notion can, in principle, be used to distinguish between different knots.

From the preceding description, it may seem that the theory of knots and links, whatever its internal beauty (a quality highly prized by mathematicians), is a rather frivolous subject. But it is not. Mathematically a knot is a simple closed curve in three-dimensional space and a link is a collection of disjoint simple closed curves; each is simple because the curve does not cross itself, and closed because the starting and end points of the curve are the same.

Simple closed curves arise in many different ways. For example, imagine a particle moving in three-dimensional space under the influence of some external force field, in such a way that after a fixed time T it returns to its original position. The path followed by the particle is a simple closed curve, provided it does not cross itself at some earlier time, and the knots and links which arise in this way contain important information. In mathematical terms these are closed trajectories of dynamical systems. Linking occurs in biological systems, for example the complicated coiling patterns of strands of DNA. In string theory, which is an attempt to model the structure of the Universe by replacing points in space-time by paths (strings) in space-time, knots and links play a role. Moffatt points out that they also occur in plasma and polymer physics.

One of the most effective ways of studying knots is through their ‘invariants’. In the simplest case this means assigning a number n_K to each knot K with the property that if n_K is different from

n_L then the knots K and L are necessarily different. Here we come across an important point. A knot may have many different representations, and a knot invariant must have the property that it is independent of the particular way we choose to represent the knot. The invariant need not be a number but may be a more sophisticated mathematical object.

The knot invariant Moffatt studies is, in his phrase, the ‘energy spectrum of the knot’. Very roughly, one can think of fluid

flowing in a small tube around the knot and then measure the energy of this flow. There are three main steps in the precise definition. Given a fixed knot K , one first constructs a knot field, which tells us the direction in which the fluid flows and is constrained to lie within a small tube enclosing the knot. This introduces some dynamics into the picture. The knot field has a well defined energy and the next step is the natural one of letting the knot and its field relax so that it assumes a configuration with minimal energy. It is necessary to specify this process quite carefully to be sure that one does not always end up with a configuration with zero energy. This could happen, for example, if one simply

B. F. Skinner (1904–1990)

BURRHUS Frederic Skinner, who died on 18 August, was the last of an earlier generation of behaviourists who dominated psychology in the United States for 25 years or more. He established his reputation with *The Behavior of Organisms* (1938), published while he was still in his early thirties. It was a remarkable book, and unlike those of Tolman, Guthrie and Hull can still be read with profit today.

Although he was preceded by the Polish neurophysiologist Konorski, Skinner was the first American psychologist to understand clearly the distinction between Pavlov’s classical conditioning, where an animal learns that one stimulus signals the occurrence of another (usually an event of some significance, like food or water), and Thorndike’s instrumental conditioning, where the delivery of food is contingent on the animal’s own behaviour. Virtually all of his own work was on instrumental or operant conditioning, conducted in a piece of equipment that everyone else came to call a Skinner box. The possibility of precise timing of events and objective recording of responses offered by this automated apparatus had a profound and (largely) beneficial effect on the study of conditioning in the laboratory. One of the perhaps unforeseen consequences was that it now became feasible to generate large amounts of data with relatively little effort. Witness Skinner’s last full-scale piece of empirical work, *Schedules of Reinforcement* (1957) — 700 large pages of experimental results, with many interesting nuggets of information and some acute observations, but mind-boggling in the ratio of undigested data to serious theoretical analysis.

Skinner abandoned the laboratory relatively early in his career: *Walden Two*, a novel about a group of people who set up a community based on sound behavioural principles of positive reinforcement, was published in 1948. It heralded a move to the public platform. If he became the most famous living psychologist, it was

not for his empirical discoveries but for his single-minded advocacy of a rather simple-minded version of behaviourism. Behaviour was to be explained by the ‘contingencies of reinforcement’ (interacting, he acknowledged, with the organism’s physical characteristics) rather than by appeal to beliefs, hopes or feelings, or even to the action of the nervous system — where that was simply inferred from the behaviour itself.

Although frequently misconstrued as denying the existence of consciousness or mental life (it was the appeal to mental states to explain behaviour that he objected to), his views, even when correctly represented, began to seem ever more naive as the ‘cognitive revolution’ swept psychology in the 1960s. Attempts to simulate human behaviour by computer programs, to provide rigorous explanations of the processes by which we perceive the world, talk, plan and act, made the simple story he continued to push seem increasingly irrelevant. And when he turned to the larger problems of society, his advocacy of behavioural engineering, based on principles of reinforcement, seemed not only naive but threatening.

Skinner will remain an important historical figure. He had a lasting effect on the study of learning and conditioning in the laboratory, and his analyses of the ways in which our own behaviour is affected (‘controlled’ in his preferred terminology) by contingencies of reinforcement contain many valuable and valid insights. Perhaps he applied them with too much enthusiasm and insufficient caution to the problems of education or psychopathology; but it is often only those who overstate and simplify their case that have any influence on their fellows.

N. J. Mackintosh

N. J. Mackintosh is in the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

scaled the field to zero. This is one of the places where Moffatt uses ideas from fluid mechanics to guide him.

The final step is to realize that there may be more than one configuration with minimal energy. The constructions start with a particular representation of the knot and if the knot is sufficiently complicated it is probable that different representations will give different configurations with minimal energy. The energy spectrum of the knot K is the sequence of numbers $0 \leq m_0 \leq m_1 \leq m_2 \dots$ given by the energy of the various configurations with minimal energy. The actual minimum energy m_0 , which on its own is insufficient to determine the knot, is the ground-state energy.

M. H. Freedman and Z.-X. He reach similar conclusions from a different, topological starting point (*Topology*, in the press). They show that if the knot has a configuration with sufficiently small energy then the knot is actually unknotted no matter how complicated it might appear to be. This result has the intuitive interpretation that it takes a lot of energy

to make fluid flow around a very complicated knot. It has the same flavour as an earlier result of J. W. Milnor (*Ann. Math.* 52, 248–257; 1950), who proved that if the total K is sufficiently small then K is unknotted. This tells us that a very complicated knot is highly curved. Freedman and He also relate the energy of the knot to other topological invariants, in particular the so-called crossing number of the knot.

The whole procedure may be reversed; the topology of knots and links may be used to obtain lower bounds on the energy of fluid flows. This is an example where topological properties, which are normally thought to be rather qualitative, force numerical constraints. Moffatt finishes his article by hinting that in the physical contexts he mentions at the beginning of his article, his energy spectrum plays an important role. But like a writer of serialized fiction, he leaves us hanging on for more. □

John D. S. Jones is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

TECTONICS

The fate of subducting slabs

Cliff Frohlich and Stephen P. Grand

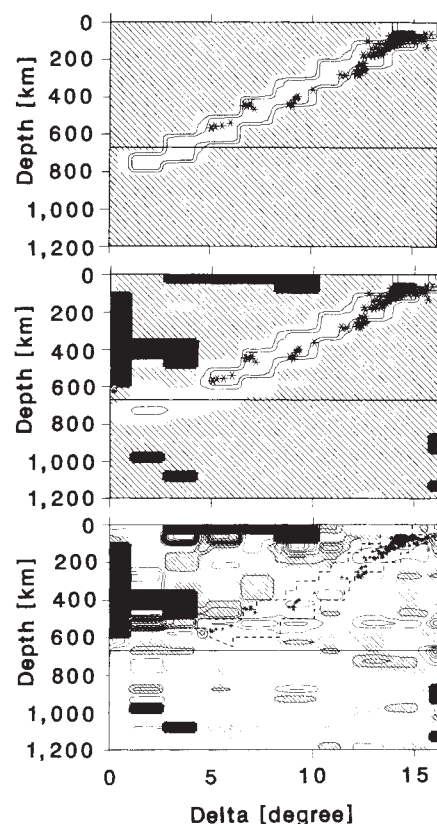
CONVERGING lithospheric plates, it is generally accepted, meet at deep ocean trenches, where one plate subducts beneath another and sinks into the mantle, accompanied by volcanism and deep earthquake activity. But once within the mantle, where does the subducted plate go? Is it absorbed by the mantle within the uppermost 670 km, the region possessing all the earthquake activity? Or does it penetrate far beneath this depth, perhaps even to the core–mantle boundary, almost 3,000 km below the surface? The questions are simple, but the answers are still a matter of controversy. Three new papers by Hua-Wei Zhou and colleagues approach them in two very different ways: residual-sphere analysis, an evaluation of anomalous travel times of individual deep earthquakes¹; and tomographic inversion of catalogued earthquake travel times to delineate seismic velocity beneath subduction zones^{2,3}. Although neither approach is new, the papers are significant because the authors use more data than previous studies, and because they are more careful to show how individual steps in each analysis affect the results.

In residual-sphere analysis, the seismologist first determines whether phases from a single deep earthquake arrive late or early at individual seismograph stations. Mapped onto a hypothetical sphere surrounding the earthquake, the data include a band of points representing phases that

have travelled through the subducting slab. The further the slab penetrates beneath the earthquake's focus, the earlier these phases arrive at the stations.

This is simple in concept but the method allows many opportunities for introducing systematic errors into the results. First, shallow near-station or deep-mantle structures produce travel-time anomalies as large as or larger than the near-source anomalies of interest. Second, many catalogued travel times are inaccurate because phases are often weak or misread. Third, travel times themselves are uncertain, as the location of the earthquake is often poorly known and affected systematically by the very structure of interest. And last, data sampling on the residual sphere is always incomplete, and often very few phases travel in the right directions to test for penetration.

The best-known previous studies, by Creager and Jordan^{4,5}, handled these difficulties by such methods as correcting for near-station structure and by statistical 'smoothing' to avoid bad or misleading data. Nevertheless, many geophysicists doubt the outcome. Zhou *et al.*¹ now construct residual spheres for no fewer than 145 deep earthquakes in the northwest Pacific. For each of 11 of these, they present figures of 10 different residual spheres showing smoothed and unsmoothed versions of the observed data as well as predicted residual spheres for four different models of mantle structure. This



Synthetic data, real tomography: how well does tomography resolve structure in the mantle? The top cross-section depicts a synthetic mantle model with higher-than-average velocity (white areas, contours are 1 per cent) in the rectangular regions that lie along a hypothetical subducting slab. If one uses this model to determine travel times at nearby seismic stations from the earthquakes (*) shown, a tomographic inversion produces the cross-section at the centre (cross-hatched areas are slow, dotted areas within 0.5 per cent of average, and black areas are unresolved by the procedure). Note how the high-velocity regions are resolved best at depths where there are earthquakes, and less well at greater depths. The bottom cross-section shows the model found by applying the same tomographic inversion method to real data beneath Japan. (Modified from ref. 2.)

allows readers to decide for themselves how well the data fit the models. For the models chosen, the authors conclude that the evidence for slab penetration is not compelling, as lower-mantle and near-station structure can account for much (but not all) of the deep-slab-like signal.

But even if the hypothesized slab signal from individual earthquakes is weak, is it somehow possible to combine the signals from all the available travel-time data? This is the approach of travel-time tomography. For earthquakes in the northwest Pacific occurring between 1964 and 1982, Zhou and Clayton² analyse more than 130,000 primary (longitudinal) and 56,000 secondary (shear) arrivals selected from the catalogue of the International Seismological Centre. The tomographic procedure divides the upper 1,200 km of the

mantle into blocks approximately 50 km deep and 200 km × 200 km in cross-section. The travel-time data are used to determine the seismic velocity within each block. Presumably, any penetration of slabs beneath 650 km would produce a group of higher-velocity blocks at this depth along the extension of the planar earthquake zone. Zhou and Clayton do find fast blocks about 800 km beneath the Kuriles, the Marianas and, in a companion paper³, Tonga. But in Japan, Izu-Bonin, the New Hebrides and the Marianas they also find fast regions spreading laterally away from the earthquake zone at depths of about 500–650 km. The authors conclude that slabs may penetrate the lower mantle in some regions, but that elsewhere they may be deflected at this 650-km discontinuity.

So where does this leave us? Unfortunately, both residual-sphere analysis and travel-time tomography suffer from a fundamental problem that occurs whenever numerous observations are filtered and fitted to models having many free parameters. In particular, a few (possibly bad) data points, or even just one, may control important details of the model — and which data are responsible may not be obvious. For the residual-sphere studies, none of the investigators^{1–4,5} checked the veracity of individual catalogued travel times by re-reading the original seismograms, a time-consuming but more convincing approach being taken for a few events by one of us (S.P.G.) and Ding⁶. With the data available at present, the slab signal is neither so obvious nor so clearly absent that it will convince believers on either side of the controversy.

Furthermore, believers in slab penetration will find it somewhat disconcerting that Zhou *et al.* can consistently resolve slabs only in the upper 650 km or so of the mantle, especially as other investigations using essentially the same data obtain different conclusions⁷. Clearly even more analysis is necessary to evaluate the resolution of the method below the depth of the earthquakes that produce the data used for the inversion. What would happen if the authors used a selective 'jackknifing' technique, restricting the inversions to data from earthquakes with focal depths in the upper 300 km? Would

they be able to resolve slabs in the intermediate depth range, 300–650 km? Zhou and Clayton do perform synthetic tests (see figure) to check the resolution of their tomographic technique, but these tests suggest resolution is much poorer beneath the deepest earthquake data sources.

Finally, the systematic problems inherent in these approaches are severe enough that it is likely that resolving the penetration controversy will require other, completely independent approaches. For example, Silver and Chan⁸ and Cormier⁹ have analysed the broadening of waveforms for phases travelling within the slab from earthquakes in the Kuriles, concluding that the slab does penetrate. In contrast, Lay and Young's similar analysis¹⁰ in the same area indicates that deep-mantle

structure, not slab penetration, causes most of the observed broadening.

If we must rule now on the fate of the subducting slab, the recent papers suggest that some form of penetration beneath 650 km occurs in some subduction zones, but not in others. This result would permit mantle mixing in some regions and not in others, possibly explaining geochemical results that seem inconsistent with ubiquitous whole-mantle mixing¹¹. If slabs do penetrate in some regions and not others, the interesting question now becomes why the regions differ mechanically. □

Cliff Frohlich is in the Institute for Geophysics, and Stephen P. Grand the Department of Geological Sciences, University of Texas, Austin, Texas 78759, USA.

MOLECULAR EVOLUTION

Old dead rats are valuable

Jared M. Diamond

THE polymerase chain reaction (PCR) has made it feasible to extract, amplify and sequence DNA from museum and archaeological specimens. In previous studies the technique has been used for taxonomic purposes by comparing DNA of extinct taxa (such as the thylacine and quagga) with that of putatively related extant taxa. W. K. Thomas, S. Paabo, F. X. Villablanca and A. C. Wilson (*J. molec. Evol.* **31**, 101–112; 1990) have now taken this approach a step further. They compare modern and historical specimens of the same extant species with respect to individual and geographical variation in mitochondrial DNA, and thereby directly measure the historical trajectory of genotype frequencies.

Thomas *et al.* used the Panamint kangaroo rat (*Dipodomys panamintinus*), a small rodent of the Californian and Nevadan deserts. They analysed mitochondrial DNA from dried skin of 43 study skins belonging to three subspecies of the rat, collected between 1911 and 1937 and preserved in the University of California (Berkeley) Museum of Vertebrate Zoology. In 1988 they collected 63 modern rats of the same subspecies from the same three sites and obtained DNA from liver. They then sequenced 225 base pairs from regions coding for two transfer RNA genes and an adjacent non-coding region.

Among the 106 individual rats, variation appears in 19 out of the 225 base positions. As would be expected for samples from three closely related populations, most of the variation involves trans-

itions, with few transversions and no length differences. Most variable sites are in one portion of the noncoding region.

The 106 rats yield 23 mitochondrial genotypes: twelve, seven and five in subspecies A, B and C respectively. Only one individual each of B and C shares the same



W. K. Thomas

Museum collections of flora and fauna may now be at the forefront of research in molecular evolution.

genotype; otherwise, each genotype is confined to a single subspecies. Both the number and the diversity of genotypes are greatest in subspecies A, the one with the widest geographical range, highest population density and hence largest population size. This is quite as expected: larger population translates into higher probabilities that a lineage will survive.

When one compares historical and modern samples of the same subspecies, the two samples of subspecies A share only two of twelve genotypes; those of B share two of six, those of C, three of five.

1. Zhou, H., Anderson, D.L. & Clayton, R.W. *J. geophys. Res.* **95**, 6799–6827 (1990).
2. Zhou, H. & Clayton, R.W. *J. geophys. Res.* **95**, 6829–6851 (1990).
3. Zhou, H. *Phys. Earth planet. Int.* **61**, 199–229 (1990).
4. Creager, K.C. & Jordan, T.H. *J. geophys. Res.* **89**, 3031–3049 (1984).
5. Creager, K.C. & Jordan, T.H. *J. geophys. Res.* **91**, 3573–3580 (1986).
6. Grand, S.P. & Ding, X. *EOS* **70**, 1322 (1989).
7. Kamiya, S. & Miyatake, T. *Geophys. Res. Lett.* **15**, 828–831 (1988).
8. Silver, P.G. & Chan, W.W. *J. geophys. Res.* **91**, 13787–13802 (1986).
9. Cormier, V.F. *J. geophys. Res.* **94**, 3006–3024 (1989).
10. Lay, T. & Young, C. *J. Geophys. Res. Lett.* **16**, 605–608 (1989).
11. Silver, P. G., Carlson, R. W. & Olson, P. A. *Rev. Earth planet. Sci.* **16**, 477–541 (1988).

At first this might seem to imply that modern populations differ from historical ones in gene frequencies, at least for subspecies A. But the temporally shared genotypes tend to be the commonest ones, whereas genotypes present only in the earlier or later sample tend to be rare ones detected in just one or two individuals. Out of the 43 museum specimens, 29 have genotypes also found in the modern specimens.

For each subspecies, genotype frequencies of the two temporal samples differ no more than would two samples randomly drawn from a single population of the observed diversity. That the two samples of subspecies A share only two genotypes (both common ones) is as one expects from the high diversity of genotypes in this large population.

Thus there is no indication of temporal shifts in genotype frequencies for any of the three subspecies. Put another way, the subspecific differences observed today were present over half a century ago.

Naturally, one must ask whether the DNA of the museum specimens has become altered with time. Numerous observations show that it has not: for two-thirds of the museum specimens, the DNA sequence is identical to that of modern specimens from the same locality but not from different localities; when the sequence is not identical, it at least fits into the appropriate clade of the phylogenetic tree; and the preponderance of transitions, as well as their preferential localization within the noncoding region, make sense.

This demonstration project will make life harder for those who are too narrow-minded to understand the scientific value of museum specimens. For example, H. T. Clifford *et al.* (*Nature* **346**, 602–604; 1990) have just offered a modest proposal to save money and improve the quality of taxonomic research, by destroying herbarium non-type specimens of plants that have recently undergone taxonomic revision (see also the responses in last week's issue, *Nature* **347**, 222–224; 1990). Yet DNA can be extracted not only from dried leaves and dried skin but also from hair, feathers and eggshells.

Old specimens constitute a vast, irreplaceable source of material for directly determining historical changes in gene frequencies, which are among the most important data in evolutionary biology. Until PCR became available, those data were lost forever as soon as the gene-bearing individual died. Now, however, museums with large, well-run collections of specimen series large enough for statistical analysis will be at the forefront of research in molecular evolution. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

The promise of gene ablation

Andrew Lumsden and David Wilkinson

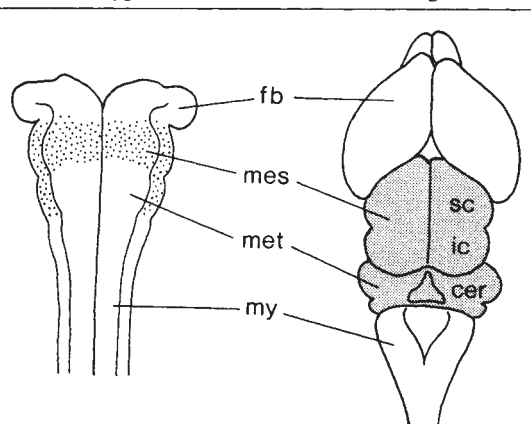
UNDERSTANDING development will depend ultimately on being able to decipher the genomic instructions for making an organism; on identifying the genes involved and elucidating the myriad interactions which regulate their function. Although the genetic control of cell differentiation is under close scrutiny, little is known about the crucial early events during the formation of the vertebrate embryo. Until now, these events have remained very much in the realm of the microsurgeon performing ablation (excision) and grafting experiments. Two reports in *Nature*¹ and *Cell*² however, describe a phenotype where pattern formation has been seriously perturbed by the 'ablation' of a gene, the proto-oncogene *int-1* (*Wnt-1*). Such an approach holds the promise of uncovering the genetic basis of development.

A number of genes display spatially restricted expression patterns in early embryos, showing that they may provide instructions during development. A direct way of testing the functions of these genes would be to inactivate them and examine the effects on the emerging pattern of the embryo. Such mutations can be generated by 'gene targeting', the substitution of a dysfunctional gene for its functional counterpart by homologous recombination. Two key advances have made it possible to inactivate genes with this technique (for a review see ref. 3) — first, the use of embryonic stem (ES) cells that can be transfected *in vitro* with constructs containing the altered gene and can contribute to the germ line when re-introduced into blastocysts; second, the development of selection and screening procedures to identify those ES cell clones in which homologous recombination has occurred.

The *int-1* proto-oncogene encodes a secreted protein that associates with extracellular matrix and thus may act as a cell–cell signal^{4,5}. Expression of *int-1* in mouse embryos is highly restricted in distribution, with transcripts being detected only in a subset of cells in the developing central nervous system (CNS)⁶. The localized expression of *int-1* at the neural-plate stage suggests that it could be involved in patterning CNS development, an idea which gains currency by the discovery that *int-1* is the homologue of a *Drosophila* segment polarity gene, *wingless* (*wg*, ref. 7). *Wingless* encodes a short-range cell–

cell signal important for the patterning of segments^{8,9}, in part by maintaining the expression of *engrailed* (*en*) in adjacent cells^{10,11}. This role cannot be extrapolated directly to vertebrates because *int-1* expression seems to be unrelated to segmentation in the CNS. But it may be significant that *wg* is also involved, without an interaction with *en*, in the specification of neuronal fate in the *Drosophila* CNS; in *wg* mutants, the phenotype of a specific neuron is altered¹².

Now Thomas and Capecchi¹ and McMahon and Bradley² report the phenotype that results from knocking out *int-1*



The expression pattern of *int-1* in the open cephalic neural plate of an 8.5-day mouse embryo (left, data from ref. 6) equates with the brain areas deleted in the brain of an *int-1* null mutant (right). The gene is expressed in the midbrain (mes) and at the edges of the anterior hindbrain (met). The superior and inferior colliculi (sc, ic) and cerebellum (cer) fail to develop when *int-1* is knocked out. The forebrain (fb) and myelencephalon (my) seem to be unaffected.

in mice. What is obvious is a substantial failure in the development of dorsal midbrain and hindbrain structures, the superior and inferior colliculi and the cerebellum being either reduced in size¹ or absent altogether^{1,2} (see figure). Ventral midbrain structures may also be deleted but the ventral anterior hindbrain (pons) appears to be unaffected. The first sign of these defects, observed by McMahon and Bradley at 9.5 days *post coitum*, is a severe diminution of the posterior two-thirds of the mesencephalon and the adjoining metencephalon.

Thomas and Capecchi, however, report a later onset of deficiencies (at 14.5 days), and also obtained a single animal which survived to adulthood. In this individual, the midbrain deletions were less severe and a small posterior part of the cerebellum persisted. Unsurprisingly, in view of its cerebellar defect, this mouse exhibited severe ataxia (loss of coordination). The reason for the discrepancy between the two studies is not clear, as both groups

disrupted the *int-1* gene in the same site, but it could be attributable to differences in the genetic background of their mice.

All the other homozygotes died shortly after birth. It is unlikely that the colliculi and the cerebellum are in themselves essential for survival, but this region of the brain is a conduit for major projections of the CNS and its deletion might be expected to bring about disruption of descending axon pathways. It also remains possible that there are undetected fatal defects in the 'engine room', the ventral myelencephalon.

There is no simple correspondence between the deficiencies and the normal sites of *int-1* expression. At 8.5 days *int-1* RNA is detected across the full medio-lateral extent of the mesencephalic neural plate and in the lateral edges of the metencephalon⁶. By 10.5 days, *int-1* is expressed in a full circle around the isthmus and extends along the dorsal midline of the diencephalon, mesencephalon, myelencephalon and spinal cord. Curiously, *int-1* RNA has disappeared from the metencephalon by this stage, yet this is the region that gives rise to much of the cerebellum.

Two explanations come to mind. On the one hand it is possible that *int-1* is critically required at the earlier stage, when the gene is expressed in the metencephalon. On the other hand, *int-1* may act following its down-regulation in the metencephalon through local cell-cell interactions, as with *wg*, beyond its immediate domain of expression. An essential, but missing, clue as to when and where this gene may act is the distribution of *int-1* protein and its receptor.

Perhaps most surprising is that the *int-1* mutants have a phenotype at all, as many of the sites of expression, including the dorsal spinal cord and diencephalon, seem to be unaffected. Many genes expressed in early vertebrate embryos are duplicated, which is perhaps in part an evolutionary provision to ensure successful development in complex organisms. When one gene fails to function, by accident or design, another family member can take over. Eleven genes are known to be

related to *int-1* (A. McMahon, personal communication) and it is possible that these genes duplicate the function of *int-1* in, for example, the dorsal spinal cord and produce a normal, regulated phenotype. Expression patterns of the other *int-1* family members have yet to be reported, but if the redundancy argument holds they should be absent from the mesencephalic-metencephalic level.

The inactivation of the *int-1* gene is an important first step towards understanding its function. But there are drawbacks to ablation experiments, whether surgical or genetic. The results are difficult to interpret both when deletions arise and, with regulation, when they do not. To understand how the null mutation leads to

the phenotype it is essential to analyse the defects fully at the cellular level and to identify genes that act downstream from *int-1*. The *en-2* gene¹³, a homologue of *Drosophila engrailed*, seems one good candidate. But the obvious defects imply that *int-1* has a crucial function in brain development, particularly in building the cerebellum, and that alone should stimulate investigation of other components of the cascade in which *int-1* is involved. □

Andrew Lumsden is in the Department of Anatomy, Guy's Hospital Medical School, London SE1 9RT, UK. David Wilkinson is in the Laboratory of Eukaryotic Molecular Genetics, National Institute for Medical Research, London NW7 1AA, UK.

MATERIALS SCIENCE

Carbon crystals wrapped up

Alan L. Mackay

AN ENTIRELY new form of pure carbon is described by Huffman and colleagues on page 354 of this issue¹. A crystalline form of football-shaped carbon shells (Fig. 1), it might find its way into novel technologies such as molecular-scale packaging or in lubricants. But its novelty is given

of crystallography from which it had earlier been excluded by researchers' concentration on those elements of symmetry that could occur in a lattice structure. Considering icosahedral shells with more than 60 asymmetric units, Caspar and Klug³ extended classical

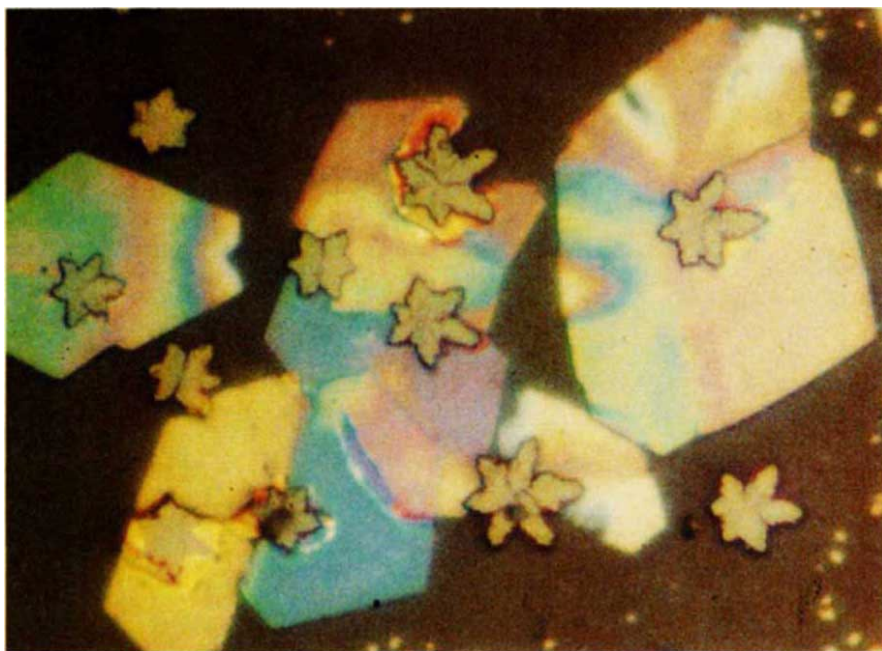


FIG. 1 Reflected-light micrograph of crystallized C₆₀ and C₇₀ fullerenes. The figure is about 50 µm across. (D. R. Huffman.)

added depth by the history of its underlying icosahedral symmetry.

In 1957, crystals of poliomyelitis virus were given to Rosalind Franklin and these enabled her, together with Finch and Klug² to deduce by X-ray diffraction, a structure for the icosahedral assembly of 60 subunits. Such icosahedral shells are now found to be an almost universal feature of spherical viruses. This work brought the icosahedron into the domain

crystallography by the concept of quasi-equivalence, in which exact identity is replaced by approximate equivalence when a hexagonal sheet is wrapped round an icosahedron.

Icosahedral symmetry also brought to Birkbeck College Richard Buckminster Fuller, the architect and engineer who had popularized the tessellations of the icosahedron in his geodesic domes. For containing space with the minimum mat-

1. Thomas, K.R. & Capecchi, M.R. *Nature* **346**, 847–850 (1990).
2. McMahon, A.P. & Bradley, A. *Cell* **62**, 1073–1085 (1990).
3. Hogan, R. & Lyons, K. *Nature* **336**, 304–305 (1988).
4. Papkoff, J. & Schryver, B. *Molec. cell. Biol.* **10**, 2723–2730 (1990).
5. Bradley, R.S. & Brown, A.M.C. *EMBO J.* **9**, 1569–1575 (1990).
6. Wilkinson, D.G., Bailes, J.A. & McMahon, A.P. *Cell* **50**, 79–88 (1987).
7. Rijsewijk, F. *et al.* *Cell* **50**, 649–657 (1987).
8. Baker, N.E. *EMBO J.* **6**, 1765–1770 (1987).
9. van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P.A. *Cell* **59**, 739–749 (1989).
10. DiNardo, S. *et al.* *Nature* **332**, 604–609 (1988).
11. Martinez-Arias, A., Baker, N.E. & Ingham, P.W. *Development* **103**, 157–170 (1988).
12. Patel, N.H., Schafer, B., Goodman, C.S. & Holmgren, R. *Genes Dev.* **3**, 890–904 (1989).
13. Davis, C.A., Noble-Topham, S.E., Rossant, J. & Joyner, A.L. *Genes Dev.* **2**, 361–371 (1988).

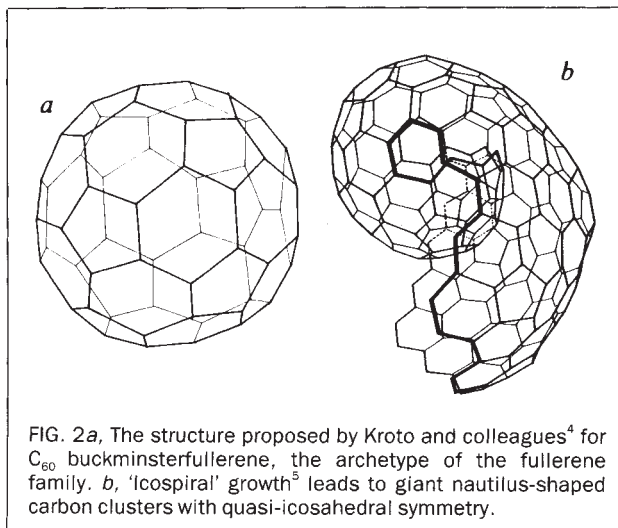


FIG. 2a, The structure proposed by Kroto and colleagues⁴ for C_{60} buckminsterfullerene, the archetype of the fullerene family. b, 'Icospiral' growth⁵ leads to giant nautilus-shaped carbon clusters with quasi-icosahedral symmetry.

erial, these domes were the best arrangement. Fuller used to ask "How much does your building weigh?" and emphasized that his could be air-lifted and might be used in space stations. Curiously, the discovery⁴ in 1985 of supposedly icosahedral molecules suggested that these miniature geodesic spheres may already exist in interstellar space.

History almost repeated with this discovery. From vaporized carbon, Kroto and colleagues⁴ produced from molecules which they showed by mass spectrometry to be C_{60} . Assuming that the sixty atoms had identical environments, they postulated that the molecules had an icosahedral shell structure and called the new polymorphic form of carbon buckminsterfullerene (Fig. 2a), even though the shape, one of the 13 archimedean semi-regular solids, had been known in antiquity and was illustrated in Kepler's works. (The shape is also universally used in footballs, although in Plato's time a dodecahedral shape had been used.) Later, Kroto and McKay⁵ suggested that the material might occur in circumstellar dust. They pointed also to larger tessellations which should be chemically possible and proposed an 'icospiral' growth mechanism (Fig. 2b) by which giant fullerene-type molecules could grow with nautilus-like structures.

In the new work, Huffman and colleagues have carried this further in producing the icosahedral C_{60} molecules in gram quantities by evaporating graphite into helium at a pressure of 100 torr and extracting C_{60} molecules with benzene. They confirm the composition using mass-spectrometry and infrared and ultraviolet spectroscopy. There is little indication of any CH impurities. They have also produced crystals (Fig. 1) which are micrometre-sized hexagonal platelets with unit cells 10.02 Å across and separated vertically by 16.36 Å, corresponding closely to a disordered simple hexagonal close packing.

The preparation of these larger quantities of the C_{60} polymorph of carbon opens

the way to the examination of more properties and many technical possibilities. The fact of crystallization indicates considerable purity. How ordered can the crystals be made? If true crystals are the stable form of C_{60} , then the atoms cannot have exactly identical surroundings and are only quasi-equivalent.

Will the compressed material be hard or soft? There is opportunity for theoretical chemists to predict properties. We should recall that the related capsule-like $C_{20}H_{20}$,

dodecahedrane, has been synthesized by Paquette⁶ and has an extraordinary stability with a melting point of $430 \pm 10^\circ\text{C}$. Elser and Haddon⁷ have predicted that C_{60} , in contrast to graphite, should be only weakly diamagnetic.

EMBRYOLOGY

Activins and induction

Chris Wylie

How does a ball of seemingly identical cells (the blastula), derived by mitotic divisions from a single fertilized egg, become differentiated into an organism of recognizably different regions and tissues? The answer to this now-standard rhetoric is, of course, that in most, if not all species, the cells are not in fact identical, but consist of at least two populations, one responding to signals from the other (see figure, overleaf). This much is known for amphibian blastulae from the now classical experiments of Pieter Nieuwkoop in the early 1970s¹. But what are the signals? Answers to this question are being marshalled through a series of elegant experiments, reported by Green and Smith on page 391 of this issue² and by others elsewhere³⁻⁷, which show that the activin class of peptide growth factors (PGFs) may be the important signal molecules in the *Xenopus* blastula.

Nieuwkoop¹ showed that when an animal cap (the normal fate of which is to form epidermis and neural tissue) is removed from the blastula, and grafted directly onto a vegetal mass, it forms recognizable mesodermal structures such as muscle and kidney: left to itself, it will form only epidermis. Furthermore, the mesoderm forms in a dorsoventral pattern imposed upon it by the orientation of the vegetal mass. These results suggest that the normal formation of mesoderm in the marginal zone *in vivo* is due to an inductive signal from the vegetal mass, which also contains a presumptive dorsoventral

Tessellated icosahedral shells are widely used in nature, by viruses for containing RNA, by vesicles and by clathrates. The fullerenes (and fullerite, as the authors have named the crystalline material) invite similar technical applications as containers for smaller molecules or radioactive isotopes, and even as micro-balloons. And given their graphite-like chemical stability, with a result of their extensive pattern of conjugated bonding, they might find use as a new kind of lubricant. □

Alan L. Mackay is in the Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK.

1. Krätschmer, W., Lamb, L.D., Fostiropoulos, K. & Huffman, D.R. *Nature* **347**, 354–358 (1990).
2. Finch, J.T. & Klug, A. *Nature* **183**, 1709–1714 (1959).
3. Caspar, D.L.D. & Klug, A. *Cold Spring Harbor Symp. quant. Biol.* **27**, 1–24 (1962).
4. Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F. & Smalley, R.E. *Nature* **318**, 162–163 (1985).
5. Kroto, H.W. & McKay, K. *Nature* **331**, 328–331 (1988).
6. Paquette, L.A. in *Strategies and Tactics in Organic Synthesis* 175–200 (Academic, 1984).
7. Elser, V. & Haddon, R.C. *Nature* **325**, 792–795 (1987).

axis. They also show that the animal cap always develops into epidermis unless diverted to form other tissues by exogenous signals.

The current excitement was generated by Jim Smith's observation that culture medium from a *Xenopus* cell line (XTC cells) caused isolated animal caps firstly to mimic the extension movements normally performed by the marginal zone during gastrulation, and secondly to form mesodermal tissues⁸. Attention quickly focused on PGFs. Purified basic FGF (fibroblast growth factor; ref. 9) and TGF β 2 (transforming growth factor; ref. 10) were reported to generate mesodermal tissues, but notably not the most dorsal ones. After an epic purification procedure, the XTC mesoderm inducing factor (XTC-MIF) was identified as a PGF (ref. 11) of relative molecular mass 24,000, and shown to induce all mesodermal tissues at nanomolar concentrations. So the stage was set for the final identification of the inducing substance. As a candidate molecule, basic FGF had the advantage of its known presence in the embryo, as did a TGF β -like molecule known as Vg1 (ref. 12).

Two major advances reported in the past two months have been, first, the identification³⁻⁵ of XTC-MIF as a homologue of mammalian activin A, a PGF previously associated with the release of follicle-stimulating hormone from the pituitary gland¹³ and with erythropoiesis¹⁴; and second, the discovery⁶ of a new cell-

line-derived inducing factor, PIF, which induces a more extensive range of tissues than seen previously, and, perhaps more important, generates extensive dorso-ventral and anterior-posterior patterning in the tissues formed.

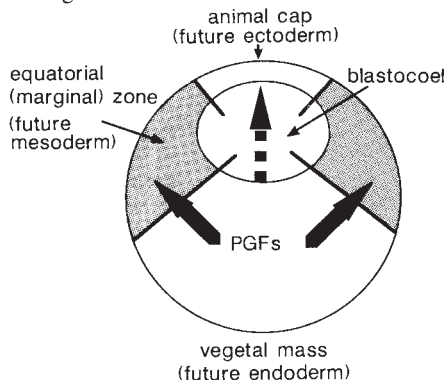
With the latest findings, Green and Smith² show that XTC-MIF is capable of generating mesoderm characteristic of different regions of the dorsoventral axis in a dose-dependent way. Disaggregated animal cap cells discriminate between remarkably small differences in the concentration of XTC-MIF. This important observation, together with the discovery of PIF, raises the possibility that a single activin A-like molecule is responsible for all the observed effects of the vegetal mass *in vivo*, acting in gradient fashion like a classical morphogen. However, the same dose of XTC-MIF induces both notochord and muscle², so things are likely to be more complicated than this, and perhaps involve several signals. Other PGFs, for example, may form part of a cascade of responses, or a balance of several PGFs may act together, or separately in time and space, to form the detailed pattern of mesoderm seen *in vivo*. Students of the control of erythropoiesis are quite used to having this many players on the field¹⁵.

Other questions will come thick and fast. What and where are the receptors? The fact that peptide growth factors are now identified as the signals means that our knowledge of these in other systems can be used to analyse their reception and transduction, as well as the early responses. How are the PGFs or their messenger RNAs positioned correctly and activated in the period of transcriptional silence before the mid-blastula stage? The events underlying these must occur during oogenesis. How do the responding cells discriminate between different concentrations of one, or several PGFs?

And what of PIF and MIF? Are they the same molecule? At first sight this seems unlikely. There are reported differences both in their chemistry⁶, and in the responses to them. Unlike MIF, PIF induces anterior neural structures (eyes and brain) in the animal cap. Current thinking suggests that this would be caused by the induction of the most dorsal and anterior mesoderm, which in turn has the property of inducing anterior neural structures in adjacent ectodermal cells. On the other hand, the biochemical differences are minor, and the differing responses could be due to differences in technique, such as the use of animal cap pieces of differing size. On balance, it seems most likely that they are the same molecule.

Lastly, is activin A the real inducer *in vivo*? Soon to be reported work from Doug Melton's laboratory^{16,17} implicates not activin A but the related activin B,

whose mRNA is found at blastula stages, whereas that of activin A appears later, during gastrulation. Furthermore, the expression pattern in early chicken embryos (where MIF and PIF have similar effects) reveals that activin B is present at exactly the right time and place to be the endogenous inducer.



Inductive signalling starts very early in *Xenopus* embryos. Cells of the vegetal mass release PGFs, starting about 4 h after fertilization. These cause cells of the marginal zone to differentiate into different types of mesoderm according to their positions along the future dorso-ventral axis; (notochord, somitic muscle, kidney and blood-forming tissue in order, from dorsal to ventral). It is thought that the blastocoel prevents induction of this response in the animal cap (see dotted line), because collapse of the blastocoel, or grafting of the animal cap directly onto the vegetal piece, causes mesodermal tissues to appear in it.

Perhaps the most satisfying aspect of these experiments is that they expose the mysterious 'plasms' of classical embryology to modern molecular analysis. They show that eggs and early embryos consist of cells with normal cellular constituents: signalling molecules and their receptors, cytoskeletal elements, adhesion molecules, transcription factors and so on. It is the fact that these elements are all harnessed to the generation of form that makes the study of early embryos so intellectually and aesthetically rewarding. □

Chris Wylie is in the Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

1. Nieuwkoop, P.D. *Adv. Morphogen.* **10**, 1-39 (1973).
2. Green, J.B.A. & Smith, J.C. *Nature* **347**, 391-394 (1990).
3. Asashima, M. et al. *Wilhelm Roux Arch. dev. Biol.* **198**, 330-335 (1988).
4. Smith, J.C. et al. *Nature* **345**, 729-731 (1990).
5. Van den Eijnden-Van Raaij, A.J.M. et al. *Nature* **345**, 732-734 (1990).
6. Sokol, S., Wong, G.G. & Melton, D.A. *Science* **249**, 561-564 (1990).
7. Albano, R.M. et al. *Development* (in the press).
8. Smith, J.C. *Development* **99**, 3-14 (1987).
9. Slack, J.M.W. et al. *Nature* **326**, 197-200 (1987).
10. Rosa, F. et al. *Science* **239**, 783-785 (1988).
11. Smith, J.C., Yaqoob, M. & Symes, K. *Development* **103**, 591-600 (1988).
12. Weeks, D.L. & Melton, D.A. *Cell* **51**, 861-867 (1987).
13. Vale, W. et al. *Nature* **321**, 776-779 (1986).
14. Eto, Y. et al. *Biochem. biophys. Res. Commun.* **142**, 1095-1103 (1987).
15. Dexter, T.M. et al. *Phil. Trans. R. Soc.* **B327**, 85-98 (1990).
16. Thomsen, G. et al. *Cell* (in the press).
17. Mitrani, E. et al. *Cell* (in the press).

Combed sound

DAEDALUS has been pondering the cocktail party effect, that uncanny ability of ours to locate and attend to one voice or sound amid a confusing babble of others all around. He now proposes to enhance this subtle spatial skill by comb filtering. A comb filter, of course, divides a sound into a large number of narrow frequency bands, and then transmits alternate bands or perhaps every third or fourth band: these days the most elaborate filters are easily implemented by digital processing. The frequency spectrum of speech has such broad features that the multiple narrow deletions of comb filtering would not remove anything critical. A comb-filtered voice would be quite intelligible and might hardly seem altered. But two voices comb-filtered out of mutual phase, the pass-bands of one coinciding with the stop-bands of the other, would now have no single frequency in common. To the ear, which is a very subtle spectrometer indeed, they would be compellingly distinct.

The first application of this cunning scheme will be to multiple telephone conversations. Most people are simply fatigued and confused by a mass of distorted voices speaking from one earpiece. While technically feasible, such conversations are not popular. But by differential comb filtering, each voice could be made as clear and distinct as at a live cocktail party. The replacement of expensive executive travel by cheap teleconferencing, long touted in vain, will take off at last.

Radio producers will also rush to exploit comb filtration. Even the best stereo reproduction is nothing like as clear as live conversation. But with each voice differently comb filtered, the protagonists in a play or discussion will seem sharply distinct from each other even on a cheap monophonic portable. A comb filter much finer than the standard musical scale could even work the same trick with music. A fine-combed soprano singer would then stand out with great clarity against her counter-combed tenor duettist or instrumental accompanist. Even the muddier choral works or more anarchic pop performances might be sharpened into defined, well resolved musical experiences.

Daedalus proposes that the different forms of electronic entertainment should all be comb filtered to mutually exclusive standards. Their combination will then be far less lethal. Mum will watch the television while Dad listens to the hi-fi, country and western streams from junior's cassette recorder and hard rock thumps through the wall from the maniacs next door. Each devotee will find it wonderfully easy to attend to his or her own choice while ignoring the rest. Domestic tranquillity, or what passes for it these days, will reign supreme.

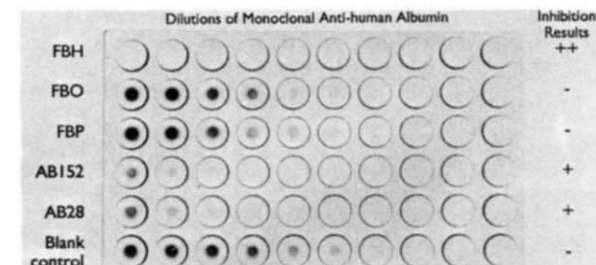
David Jones

Blood in ancient human bone

SIR—The detection of blood proteins of human or animal origin in ancient material is extremely important: it will allow identification of the species-origin of small bone fragments found on ancient archaeological sites, and therefore help reconstruct ancient ritual, dietary and domestic behaviour. It may also provide a dependable means of determining the species-origin of blood stains on stone artefacts

years. Previous researchers^{1,2} failed to attain the specificity and high sensitivity that we have achieved using ELISA and monoclonal antibodies. Although used on blood stains in forensic studies³, this technique has been applied only once in archaeology — to pigments in Australian rock-art⁴. ELISA improves the sensitivity of detection and the use of monoclonal instead of polyclonal antibodies greatly increases the specificity, reducing the risks of unsuspected proteins cross-reacting and giving false results.

Our technique offers an easy, cheap and safe method of detecting extremely small amounts (10 ng) of protein while guaranteeing the maximum specificity attainable. The method will complement the emerging techniques of nucleic-acid analysis, particularly the use of the polymerase chain reaction (PCR) to amplify DNA sequences from ancient human bone⁵. We are currently developing monoclonal antibodies against non-human species likely to be encountered on archaeological sites.



Results obtained with fresh human (FBH), Saxon ancient human (AB 152 and AB 28), fresh ox (FBO) and pig (FBP) bone extracts when tested against doubling dilutions of monoclonal anti-human albumin. The diminished colour development seen with human (+ + +) as opposed to animal (-) extracts denotes the presence of human albumin. Details of methods available from C.C. on request.

and thus permit interpretation of hunting and other practices.

To identify the blood protein albumin, we prepared extracts from the bones of 12 young adults from the Berensfield Early Saxon site (circa AD 450–600) and from two adults from mediaeval Chichester (circa AD 1100–1400), both sites being in Britain. We tested our samples for the presence of human albumin by an inhibition enzyme-linked immunosorbent assay (ELISA) using a monoclonal antibody (see figure). Both mediaeval subjects gave positive results, and 9 of the 12 Saxon individuals clearly showed the presence of human albumin, despite poor skeletal preservation. Extracts from fresh animal (ox, pig or sheep) and fresh human bone were included as negative and positive controls, respectively.

Our results show that the blood protein albumin survives for more than 1,000

everywhere (at temperatures where condensation occurs in preference to ice nucleation) to initiate cloud formation when the relative humidity H achieves a value of about 100%. The introduction of further, highly effective CCN can, however, increase the droplet concentration N with a corresponding reduction in droplet equivalent radius r and an associated increase in cloud albedo. To a first approximation, $r/r_0 = (N/N_0)^{-1/3}$, where r_0 and N_0 are the initial values of droplet equivalent radius and droplet concentration, respectively.

The number concentrations of droplets in clouds over land are characteristically several hundred per cubic centimetre, whereas over the oceans they are about an order of magnitude less. I take $N_0 = 50 \text{ cm}^{-3}$ as a typical maritime droplet concentration. From the above equation, if the concentration of droplets in maritime clouds (low-level stratus) is doubled to $N = 100 \text{ cm}^{-3}$, the cloud droplet equivalent radius r would be decreased by about 25%, which is more than sufficient to offset the warming effect of a doubling in the concentration of carbon dioxide.

I assume first that the CCN particles advertently introduced into the atmosphere are each of mass $m = 10^{-12} \text{ g}$ (radius, $0.6 \mu\text{m}$), and that they are present in concentrations of 100 cm^{-3} over the entire area of the Earth to an altitude H of 1 km. Thus the total mass of these CCN per unit area of the Earth's surface would be $M = H N m = 100 \text{ kg km}^{-2}$.

The average longevity τ of these particles in the atmosphere is likely to be about one day. To maintain the stipulated enhancement in N , CCN need to be disseminated into the atmosphere at about $4 \text{ kg km}^{-2} \text{ h}^{-1}$. If m were $= 10^{-13} \text{ g}$ (radius $\sim 0.3 \mu\text{m}$), the dissemination rate required would be about $0.4 \text{ kg km}^{-2} \text{ h}^{-1}$.

A discussion of the technology required for the generation of CCN on the required scale is beyond the scope of this correspondence. Note, however, that these dissemination rates are modest, and that the ocean surface provides an excellent site for the production (by bubble bursting, for example) of very efficient hygroscopic condensation nuclei in the form of droplets of sea water, the primary constituent of which is NaCl. In the radius range considered, such particles will be more efficient as CCN (that is, will be activated at lower supersaturations) than those originating from dimethylsulphide, which are believed³ to be the main source of CCN over the oceans. The above dissemination rates correspond to deposition rates of sea water into the oceans of about 40 and $4 \mu\text{m yr}^{-1}$, respectively.

It seems feasible to conduct an experiment in which CCN are introduced in a controlled manner into marine stratus, and in-cloud microphysical measurements and above-cloud radiative observations

Control of global warming?

SIR—Is it quantitatively feasible to introduce in a controlled manner into the atmosphere particulate material, to change the characteristics of natural low-level clouds in such a way as to inhibit or neutralize global warming? The idea that small changes in cloud cover (or cloud properties) can produce temperature changes of a few degrees Celsius comparable with those associated with enhanced long-wave absorption accompanying increased carbon dioxide concentrations is consistent with recent calculations¹. These showed that climate changes equivalent to those produced by doubling the

carbon dioxide in the atmosphere can be produced by relatively small changes in the extent of cloud cover C , the liquid-water path L , and the droplet equivalent radius r in the clouds. Specifically, it was predicted that the warming produced by the carbon dioxide content can be balanced by increases of 15–20% in C and 20–35% in L , and by decreases of 15–20% in r .

In the lower atmosphere there is an ample supply of natural cloud condensation nuclei (CCN), which act as centres on which cloud droplets are nucleated², in the sense that sufficient concentrations exist

C. CATTANEO*
K. GELSTHORPE†
P. PHILLIPS*
R.J. SOKOL†

*Department of Archaeology and Prehistory,
University of Sheffield,
Sheffield S10 2TN, UK

†Regional Blood Transfusion Centre,
Longley Lane, Sheffield S5 7JN, UK

- Lowenstein, J.M. *Phil. Trans. R. Soc.* **B292**, 143–149 (1981).
- Ascenzi, A., Brunori, M., Citro, G. & Zito, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7170–7172 (1985).
- Fletcher, S.M., Dolton, P. & Harris-Smith, P.W. *J. forensic Sci.* **29**, 67–74 (1984).
- Loy, T.H. *et al. Antiquity* **64**, 110–116 (1990).
- Hagelberg, E., Sykes, B. & Hedges, R. *Nature* **342**, 485 (1989).

made. It would also be possible to assess these ideas by incorporating them into an established cloud model, which treats the salient microphysical processes explicitly. The most propitious situation for effecting the proposed radiative changes would involve fields of stable boundary-layer clouds with low droplet concentrations, occurring under highs over the tropical oceans.

Increasing the droplet concentration in clouds is likely to enhance their colloidal stability, making the coalescence process that leads to rainfall more difficult to initiate and thereby, in some situations, pro-

longing the lifetime of the clouds. The introduction of ice-forming nuclei into colder clouds could increase or decrease their lifetimes, depending on the circumstances.

JOHN LATHAM

Department of Pure and Applied Physics,
University of Manchester Institute of
Science and Technology,
PO Box 88, Manchester M60 1QD, UK

1. Slingo, A. *Nature* **343**, 49–51 (1990).
2. Pruppacher, H.R. & Klett, D. *Microphysics of Clouds and Precipitation* (Reidel, Dordrecht, 1978).
3. Charlson, R.J., Lovelock, J.E., Andreae, M.O. & Warren, S.G. *Nature* **326**, 655–661 (1987).

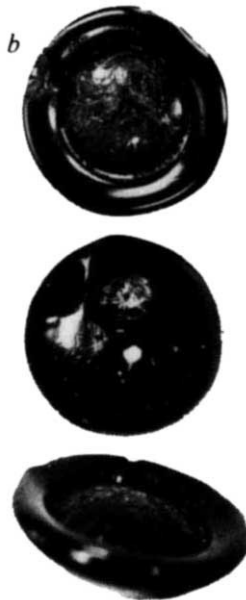
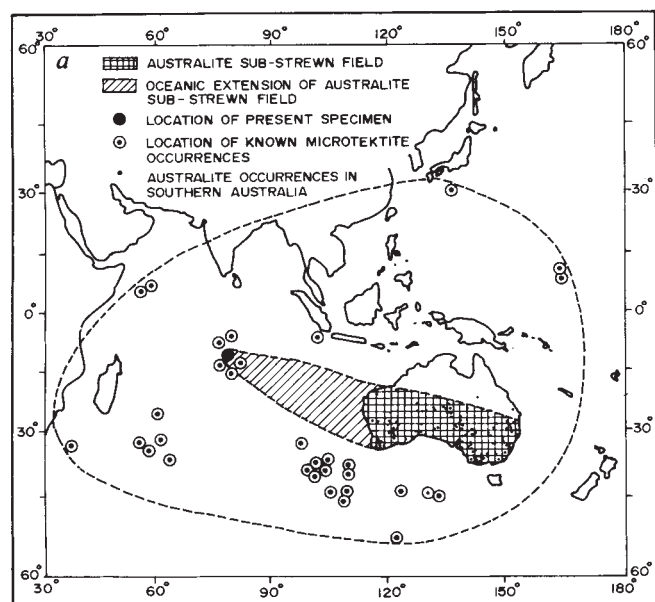
Tektites far and wide

SIR—Meteorite impacts on terrestrial rocks produce glassy ejecta called tektites¹. These have been found strewn in four fields, of which one in Australasia is both the youngest (0.7 Myr) and the largest (covering about 1/10 of the Earth's surface²; *a* in the figure). Flanged australite buttons (aerodynamically ablated tektites) have so far been found only in southern Australia, never (although expected) in the ocean. We have recovered a flanged australite button from the surficial sediments of the Central Indian Basin (12° 37' S; 78° 30' E), at a water depth of 5,300 m.

Our tektite (*b* in the figure) measures 12.054×4.092 mm, with a flange width of 2.352 mm, and weighs 0.6633 g. It is similar to the perfectly formed and well preserved Port Campbell specimens³. The specific gravity of the tektite is 2.508 and its refractive index is 1.528, corresponding to a silica content⁴ of 66 per cent. The specimen is smaller than typical flanged

buttons from southern Australia, and its density is greater than average for australites. By extrapolating the curvature of its core and estimating a volume loss of about 62 per cent during ablation, we deduce the pre-ablation shape of the tektite to have been a sphere of 11 mm diameter. The tektite has a vitreous lustre, and the excellent preservation of the fine markings of the schlieren structures on the core exclude long-distance transport. The sedimentation rate in this area is 2 m Myr^{-1} (ref. 5), so that, given the 0.7-Myr age of the Australasian tektite field, this tektite should have been buried under 1.4 m of sediment. Its near-surface retention could be due to bioturbation and/or removal of sediments by bottom currents; alternatively, the Australasian tektites might have fallen more recently⁶.

Our finding signifies a large westward extension of the australite sub-strewn field into the Indian Ocean, increasing its area by about 4.3 million km^2 (*a* in the figure).



a, The Australasian tektite strewn field (from ref. 2; australite distribution in southern Australia is from ref. 7) and the oceanic extension suggested by our finding. *b*, Top: posterior view of the flanged australite button, showing well-preserved, schlieren structures on the core and smooth texture on the flange. Middle: anterior side, showing a bubble pit in the top half. A thin circular groove runs along this side; further ablation would have led to flange shedding along the groove, resulting in the formation of a lens. Bottom: oblique view from posterior side, showing curvature of the core. The diameter of the tektite is 12.054 mm.

The australites in southern Australia occur in two sub-parallel lines representing their original lines of infall⁷. The distance between these two lines relative to the oceanic extension of the australite sub-strewn field suggests that this tektite could be a part of one or more such lines of infall, the existence of which has hitherto been undetected in the oceans because of poor sampling. This finding does not conform to Von Koenigswald's⁸ observation of latitudinal morphological similarities between Australasian tektites, and also Ford's⁹ proposed southeasterly distribution of Australasian tektites and Chapman's¹⁰ suggestion that aerodynamically ablated australites do not occur north of 23° S.

M. SHYAM PRASAD

P. S. RAO

National Institute of Oceanography,
Dona Paula,
Goa 403004, India

1. Taylor, S.R. *Earth Sci. Rev.* **9**, 101–123 (1973).
2. Glass, B.P., Swincki, M.B. & Zwart, P.A. *Proc. Lunar planet. Sci. Conf.* **10**, 2535–2545 (1979).
3. Baker, G. in *Tektites* (ed. O'Keefe, J.A.) 1–24 (University of Chicago Press, 1963).
4. O'Keefe, J.A. *Tektites and their Origin* (Elsevier, Amsterdam, 1976).
5. Udintsev, G.B. *Geological-Geophysical Atlas of the Indian Ocean* (Pergamon, New York, 1975).
6. Chalmers, R.O., Henderson, E.P. & Mason, B. *Smithson. Contrib. Earth Sci.* **17**, 1–46 (1976).
7. McColl, B.H. & Williams, G.E. *Nature* **226**, 154–155 (1970).
8. Von Koenigswald, G.H.R. *Proc. K. Ned. Akad. Wetensch.* **70**, 104–112 (1967).
9. Ford, R.J. *Austr. J. Earth Sci.* **35**, 483–490 (1988).
10. Chapman, D.R.J. *J. geophys. Res.* **76**, 6309–6338 (1971).

More light on PCR contamination

SIR—In our recent Scientific Correspondence¹ we showed that ultraviolet light was a potent inactivator of a 750-base pair segment of DNA when it contaminated reagents. All reagents, including *Taq* polymerase, could be decontaminated but, for obvious reasons, contamination of the template DNA cannot be eliminated. Despite a 100,000-fold inactivation of contaminant, the integrity of all the reagents was preserved to the extent that a normal intensity of amplified product could be seen in a standard 30-cycle polymerase chain reaction of human genomic DNA. Cimino *et al.*² did not appreciate some of the above points. They presented a theoretical argument and reported an experiment indicating that a 121-bp segment is markedly less susceptible to ultraviolet inactivation than an unrelated 500-bp segment.

We have performed experiments which support and extend this finding. Sensitivity to ultraviolet varies enormously ($> 10^3$) in the 11 segments examined. Sequence is important as well as size as a 151-bp segment was at least 25-fold more sensitive than a 430-bp segment. Two main

practical conclusions emerge from our studies. First, whenever feasible, large segments of DNA should be amplified as we found five out of five DNA segments > 700 bp to be highly susceptible to ultraviolet inactivation in contrast to only one of four segments < 250 bp. Second, ultraviolet inactivation is much less effective in eliminating dried DNA, suggesting that decontamination of laboratory equipment may require the aid of a photosensitizer or even a completely different approach.

GOBINDA SARKAR
STEVE SOMMER

Department of Biochemistry and
Molecular Biology,
Mayo Clinic/Foundation,
Rochester,
Minnesota 55905, USA

1. Sarkar, G. & Sommer, S.S. *Nature* **343**, 27 (1990).
2. Cimino, G.D. et al. *Nature* **345**, 773–774 (1990).

Repetitive orbital damage

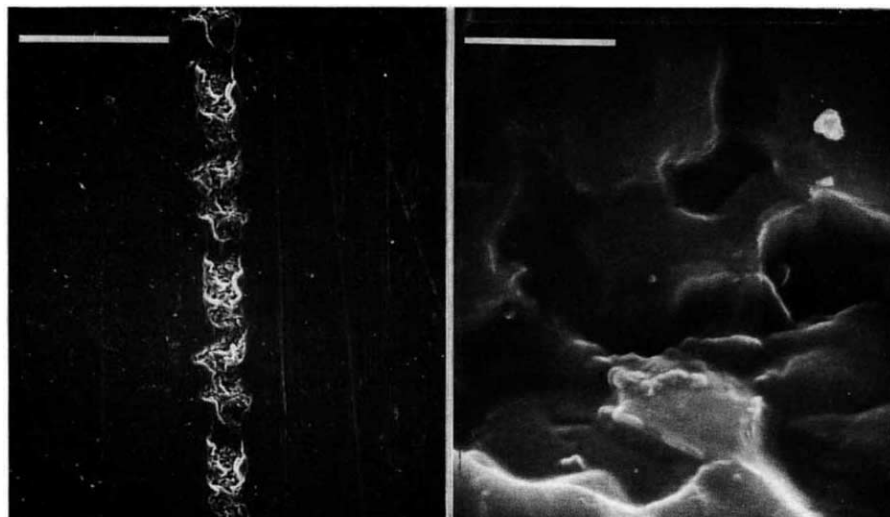
SIR—In the course of a preliminary routine search for micrometeoroid strikes on an experiment tray on a low-Earth orbit satellite we observed several long strings of repetitive indentations which, on detailed examination in the electron microscope, demonstrate similarities of structure down to an extremely fine scale. The experiment panel was flown for 5.8 years in circular Earth orbit at altitudes of

sample cut from the same piece of the frame containing the under (unexposed) side of the frame. The material of both samples was structural aluminium (2024-T3) clad with commercially pure aluminium. After the mounted specimens were vacuum-coated with approximately 20 nm of gold, we examined them in a Hitachi CS-800 field-emission scanning electron microscope at 15 kV.

The left-hand side of the figure shows a short segment from a string about 6 mm long, incorporating five individual units. From the top down, units, 1, 3 and 5 are the same and units 2 and 4 are the same. This sequence continues perfectly throughout almost all the 6-mm length, except that at the ends the indentations become less distinct. The right-hand side of the figure is a detail from unit 2.

We do not know the source of the indentations. We have observed 10 separate strings on the exposed surface examined, all having a level of replication quality similar to the string shown here. Of the ten strings examined, eight have only one kind of repeating unit, while two have an alternating sequence such as those shown here. We have not found any string that bears any resemblance to any other, but all individual units are 20–60 µm in size.

We suggest that the features observed were not produced during fabrication or handling of the aluminium frame. Because the experiment tray was either protected in a clean room or in orbit from the time



Left, scanning electron micrograph of five units of a 6-mm-long string. Scale line is 75 µm long. Up is away from the Earth, left is the direction of velocity of the experiment tray. Right, portion of the lower left extension of unit 2 (counting from the top). Scale, 1.76 µm.

333 to 481 km. The plane of orbit of the vehicle was approximately 5° from the ecliptic. The orientation of the normal to our panel ranged during each passage around the Earth through an arc from 87.25° to the plane of the ecliptic to 77.25° in the opposite direction.

We cut a 20 × 15 mm sample of the outer exposed surface from the aluminium frame and mounted it next to a 20 × 8 mm

NASA mounted it on the satellite until it was returned to us, the evidence indicates that the observed indentations were likely to have been created during the 5.8-year mission.

D. K. FELBECK
W. H. DURRANT

University of Michigan,
Ann Arbor,
Michigan 48109-2125, USA

Origin of vpx in lentiviruses

SIR—Primate lentiviruses have been classified into four groups on the basis of similarity of their Pol gene products, as discussed by Desrosiers' News and Views¹. All these viruses have five genes that are essential, in the cases examined, for propagation *in vitro*: the structural genes *gag*, *pol* and *env*, and the regulatory genes *tat* and *rev*. Five other genes, not essential *in vitro*, are found in various combinations: *vif* and *nef* are invariably present but *vpu* is found only in the HIV1/SIV_{CPZ} group; *vpr* is found in all viruses except those of the SIV_{AGM} group; and *vpx* is found in only two groups — SIV_{AGM} and HIV2/SIV_{SMN}/SIV_{MAC}.

We suggest that the *vpx* gene in the HIV2/SIV_{SMN}/SIV_{MAC} group arose by duplication of the *vpr* gene. The strongest support for this proposal lies in the alignment of the Vpr and Vpx proteins from various prototype isolates shown in Fig. 1. All the Vpr and Vpx proteins share regions of similarity with each other, suggesting critical residues whose mutation may help to elucidate their functions.

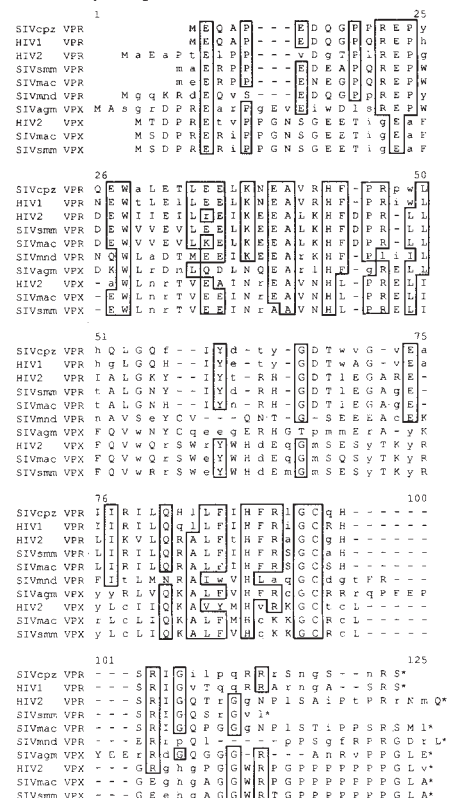


FIG. 1. Comparison of the deduced amino-acid sequences of *vpx* and *vpr* genes from primate lentiviruses. The following isolates were used in the comparison: HIV1_{BRU}, SIV_{CPZ}, HIV2_{ROD}, SIV_{SMN}₄, SIV_{MAC}_{MM142}, SIV_{MND}_{GB1}, SIV_{AGM}_{TYO-1}. Sequences obtained from the Human Retroviruses and AIDS Database^{3,4}. Capital letters, conserved amino acids between isolates; amino acids boxed when identical in six or more cases. Dashes, gaps introduced in the sequences to allow optimal alignment. Sequences aligned both by eye and with the programs AMPS⁵ and HOMED⁶.

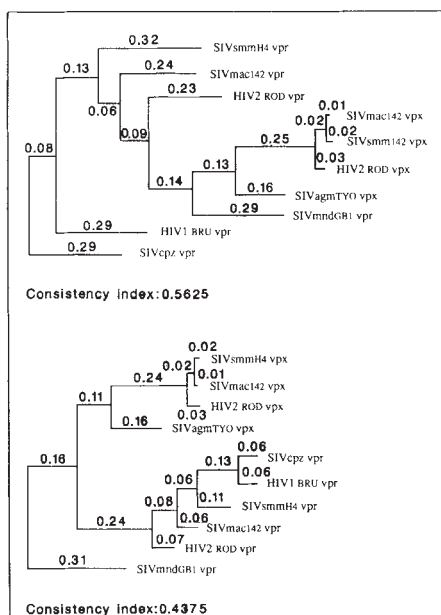


FIG. 2. Minimum-length unrooted evolutionary trees inferred from the protein sequence alignment in FIG. 1. The two different trees found were calculated from the protein sequence parsimony method (PROTPARS) using the PHYLIP package⁷. Lengths of horizontal lines are proportional to the fewest nucleotide substitutions required to modify the codons to generate the observed amino-acid variation. The figure above each line is the ratio of the branch length to the total number of sites. The length of the vertical lines is for clarity only.

Further support for the duplication lies in the fact that the *vpx* and *vpr* genes are adjacent to each other when both are present. A precedent exists for a substantial viable duplication in a retrovirus²: the simian AIDS retrovirus (SRV-1) has been shown to have a duplicated protease gene in which the duplicated regions also lie in tandem.

To compare the relationship of *vpr* and *vpx* genes further, we constructed evolutionary trees (Fig. 2) which show that a single ancestral gene probably evolved into all the *vpx* or *vpr* genes. The *vpx* genes may have diverged sufficiently so that they now perform a new function, explaining the retention of the duplication. Consistent with this, the constraints on *vpx* seem, from the branch lengths shown in Fig. 2, to be greater than those on *vpr*.

The SIV_{AGM} Vpx protein aligns well with all Vpr proteins and sometimes better with the Vpr of a particular virus than with the respective Vpx. We suggest that the presence of a *vpr* or *vpr*-like gene is a

common and distinguishing feature of the lentiviruses found in primates. We therefore propose that the classification of the SIV_{AGM} *vpx* gene be reconsidered.

MICHAEL TRISTEM
CRAIG MARSHALL
ABRAHAM KARPAS
JURAJ PETRIK
FERGAL HILL

Department of Haematology,
University of Cambridge Clinical School,
Hills Road, Cambridge CB2 2QL, UK

Helium increase

SIR—The Loma Prieta earthquake of 17 October 1989 was not preceded by any clear precursor from seismic, strain or geodetic measurements¹. Yet, the concentration of helium gas in the soil monitoring station in Stone Canyon on the San Andreas fault, 60 km southeast of the 7.1 magnitude epicentre showed an unprecedented increase one day before the earthquake. Soil-gas helium, collected from a 2-m-deep probe, consistently ranged from 5,160 to 5,440 parts per thousand million (p.p.b.) at the Stone Canyon site in central California (Fig. 1). By contrast to my previous observations of helium decreases preceding earthquakes², the sample for 16 October 1989, collected at 16.30 UT showed a helium concentration of 5,560 p.p.b. greater than any previously collected sample.

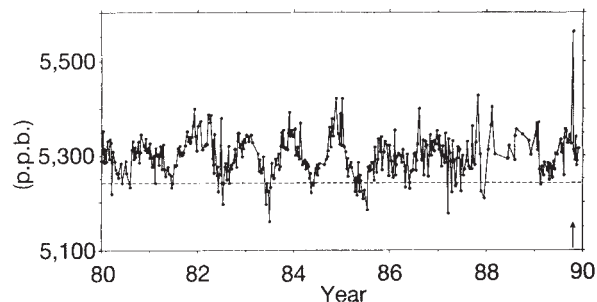
The concentration was nearly twice as high from an air baseline than any previously observed during 10 years of monitoring at this site^{2,3}. Neither of the two next closest helium stations to the epicentre, 100 km southeast on the San Andreas fault at Parkfield and 40 km east on the Calaveras fault, showed any increase in helium.

Compared to past earthquakes along other segments of the San Andreas fault, the Loma Prieta event had some unusual characteristics: the epicentre was deep, at 17.6 km, with the slippage occurring on a 70° inclined plane¹ (rather than the common near-vertical faults of the San Andreas fault); there was more than 1 m of vertical displacement in addition to the right-lateral slippage⁴ (which itself did not leave a clear surface expression); and it was the largest event (magnitude 7.1) to occur in the area since the 1906 San Francisco earthquake.

The reasons for the helium variation can be derived from experimental and empirical evidence. The release of noble gases from mineral grains before rock failure has been documented in the laboratory^{5,6} and in the field⁷. The occurrence of the vertical displacement implies

that the fault planes had sufficient permeability for rapid vertical gas transport. Gas-flow rates of the order of kilometres per hour would be required to explain the Loma Prieta helium observations if the gas was released from the deep hypocentre. Such a rate has been observed through a 5 to 9 km zone at Mount St Helens⁸, although the saturated pores of a fault zone may impede the flow somewhat. However, there is no restriction that the area of eventual rock failure must be the actual source of gas release. Helium may have been collecting in a structural trap defined by the suggested intersection⁴ of the 70° fault plane and a near-vertical plane typical of the San Andreas fault. Gas could be released from shallower zones in response to pressure gradients induced by changes in stress.

I do not suggest that any single type of observation would be a successful predictor of seismic activity. But when the total body of observational evidence is examined, no one technique, including seismic monitoring, has been consistent in providing precursory signals. Clearly, variations in the natural system in both time and space suggest that a technique which successfully records a precursory



Soil-gas helium concentrations for the Stone Canyon sample location for 1980–1990. Arrow, Loma Prieta earthquake; dashed line, helium concentration in air at 5,240 p.p.b. used as a reference baseline. Seasonal cyclicity from meteorologic influences is apparent, but the maximum concentrations did not exceed 5,440 p.p.b. Estimation of soil retention-times for helium is available from gas injection experiments (G.M.R. unpublished data). If the soil-gas concentration has not increased more than 1,000 p.p.b. from an episodic release, the sample would not show a residual of the increase 1 week later.

phenomenon on one occasion may not be successful the next. Effective earthquake prediction will have to involve many monitoring techniques

G. M. REIMER

MS 939, US Geological Survey,
Denver Federal Center,
Denver,
Colorado 80225, USA

- Desrosiers, R. *Nature* **345**, 288–289 (1990).
- Power, M. D. *et al.* *Science* **231**, 1567–1572 (1986).
- Myers, G. *et al.* *Human Retroviruses and AIDS* (Los Alamos Lab., Los Alamos, USA, 1989).
- Myers, G., Berzofsky, J. A., Rabson, A. B., Smith, T. F. & Wong-Staal, F. *Human Retroviruses and AIDS* (Los Alamos Lab., Los Alamos, USA, 1990).
- Barton, G. J. & Sternburg, M. J. *E. J. molec. Biol.* **198**, 327–337 (1987).
- Stockwell, P. A. *Trends biochem. Sci.* **13**, 322–323 (1988).
- Felsenstein, J. A. *Rev. Genet.* **22**, 521–565 (1988).

- US Geological Survey Staff, *Science* **247**, 286–293 (1990).
- Reimer, G.M. *Pure appl. Geophys.* **122**, 369–375 (1985).
- Reimer, G.M. *Abstr. 28th Int. Geol. Cong.* **2**, 685–686 (Washington, DC, 1989).
- Plafker, G. & Galloway, J.P. (eds) *US Geol. Surv. Circ.* **1045** (1990).
- Honda, M., Kurita, K., Hamano, Y. & Ozima, M. *Earth planet. Sci. Lett.* **59**, 429–436 (1982).
- Holub, R.F. & Brady, B.T. *J. geophys. Res.* **86**, 1776–1784 (1981).
- Barsukov, V.L., Serebrennikov, V.S., Varshal, G.M. & Garanin, A.V. *Geochem. Internat.* **16**, 1–13 (1979).
- Weaver, C.S., Zollweg, J.E. & Malone, S.D. *Science* **221**, 1391–1394 (1983).

Ruling the waves

R.V. Jones

Metres to Microwaves. By E.B. Callick. *Peter Peregrinus Ltd: 1990. Pp. 240. £38.*

"DEPEND upon it, Sir, when a man knows that he is to be hanged in a fortnight", said Samuel Johnson, "it concentrates his mind wonderfully" — and this was the prospect that loomed before us in Britain with the rise of the Nazi menace in the 1930s. Above all, the threat of bombing by the *Luftwaffe*, and the inability of our existing air defences to counter it, drew serving officers and scientists together in a joint effort to devise and perfect new methods. The outstanding result, of course, was the radar system that turned the balance in the Battle of Britain.

As regards radar, that battle was fought at metric wavelengths because at the time only at these wavelengths could sufficient power be generated to produce detectable echoes from such targets as single aircraft at ranges sufficient for fighters to be scrambled to intercept the approaching bombers. In *Metres to Microwaves*, E.B. Callick gives a uniquely informative account of the development of the "active components" — transmitter and receiver valves, cathode-ray tubes and duplexers for a common transmitter-and-receiver working from a single antenna — that were vital to the functioning of radar stations. This aspect has rarely been treated in earlier books. Popular attention has inevitably concentrated on the operational aspects of the Battle of Britain, but it should be remembered that there were great difficulties to be overcome in the nascent technology of electronics, and it is good to read of how successes were achieved.

Indeed, it is only right that these successes should be recorded alongside their operational counterparts, for they were no less vital. Even as late as 1937 I can recall sitting in front of a flickering cathode-ray tube at Bawdsey, surrounded by 'bread boards' and wondering whether such 'hook-ups' could possibly be converted into a reliable working system in time for the prospective battle. It is to the indelible credit of all concerned that the system was indeed ready for 1940.

Effective though the system was, it was also evident that something better would be needed for night battles and for other purposes to which radar techniques could be applied. In particular, new methods were needed to generate high power at centimetric, rather than metric, wavelengths because of the benefits that would result from the sharper beams that could thereby be created. Here was a case of necessity calling for invention, and Callick gives a lucidly detailed account of how the

cavity magnetron was developed, notably by J.T. Randall and H.A.H. Boot in Birmingham, with the collaboration of R.W. Sutton in Bristol, E.C.S. Megaw at GEC and researchers at the Clarendon Laboratory in Oxford. The contribution of H. Gutton in Paris is also recognized: Gutton's use of the barium oxide-coated cathode, thanks to its low electronic 'work function', gave a tenfold increase in

Callick is eminently qualified to write this account, for he himself was continuously involved, in a career that included work in industry at GEC and EMI, in government service at Farnborough, Malvern and Baldock, and as chief engineer of the English Electric Valve Company. Among his predecessors as secretary of the Committee on Valve Development (CVD) in the British Admiralty was J.H.E. Griffiths, who discovered ferromagnetic resonance, and whose enormous fund of tact was on one occasion strained by a telephone call from an earnest signals brigadier in the War Office who, finding that magnetrons, with which *Metres to Microwaves* is much concerned, were coming into wide use, heard that there was

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A scientific officer, E.R. Tobitt, adjusting the turning gear on the aerial of the first British automatic-tracking gun-laying radar equipment in 1944. (Courtesy of D. H. Tomlin).

output power, with secondary emission providing 99 per cent of the operating current. Similarly, the invention by J.L. Sayers of "strapping" between magnetron cavities gave a further manifold increase in output.

As well as covering the important technical details, Callick gives an account of the administrative background and provides valuable appreciations summarizing the key achievements made in the fields that he surveys. Many of the named contributors were men whose main pre-war interests lay outside electronics, but who were able to transfer their specialist experiences — including Alan Hodgkin, who has written the foreword, J.W.S. Pringle, W.B. Lewis, P.I. Dee and H.W.B. Skinner who, at the suggestion of M.L. Oliphant, converted the ancient 'cat's whisker' and crystal into an effective detector for microwaves. This success was one of the factors that led to the postwar invention of the transistor.

another device called a cyclotron. "I say, Griffiths" said the brigadier, "what's a cyclotron?" Griffiths did his best to explain but the brigadier was little wiser — "Couldn't you just bring one over to the War Office so I could have it on my desk to look at?"

If humour may be permitted into a review of an eminently serious book, there was much of it among those who contributed to the developments that the book describes. New recruits to Bawdsey were, for example, taken on an introductory tour of the various makeshift huts in which the several components of the radar system were being developed. Two particularly gullible new scientific officers were cautioned before they entered the transmitter hut and asked whether they were wearing their "protectors". When they expressed surprise, they were told that everyone had to wear these because, with more than 10,000 volts on the transmitting anodes, X-rays were emitted. As

these were known to produce sterility, the Air Ministry had designed lead protectors to be worn around the relevant parts of the male anatomy, and the recruits had better go back to the Bawdsey store and requisition pairs for themselves. To add conviction, some of those in the know carried metal objects in their trouser pockets which could be jangled on appropriate occasions. The recruits were so convinced that they in turn convinced the store-keeper, who actually indented for the hypothetical protectors from RAF central stores. It was only after the whole station had laughed at the joke that the head of the transmitter hut, ruminating on the incident, decided his theory was so plausible that he had better check it. Wrapping photographic plates in opaque paper, he found that they were indeed rapidly blackened by the hitherto unsuspected X-rays.

It was an exhilarating time, in which innovative brilliance, frantic work, humour and indeed tragedy (including the loss of A.D. Blumlein and others who died in trials of experimental equipment) closely intermingled. *Metres to Microwaves* tells the solid story of the technical achievements which gave Britain and its allies an overwhelming advantage in the electronic war. □

R.V. Jones is at 8 Queen's Terrace, Aberdeen AB1 1XL, UK.

AIDS TARGETED INFORMATION/ATIN

TARGETS YOUR LITERATURE SEARCH

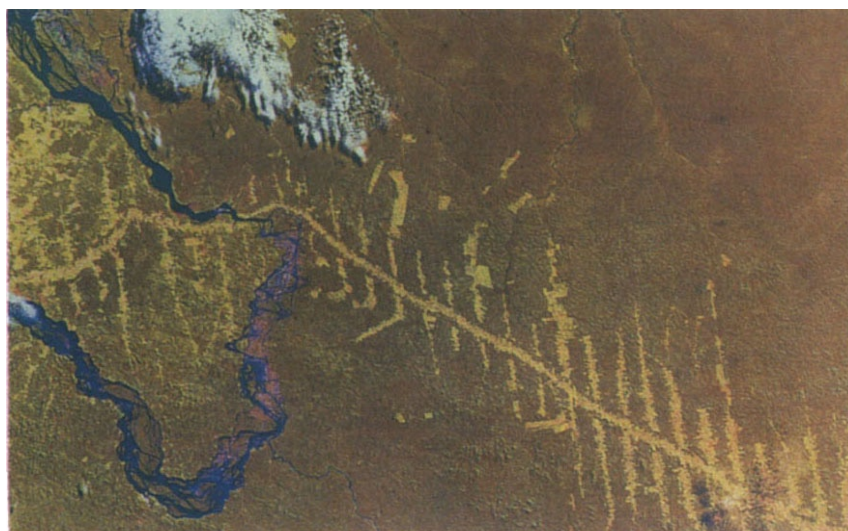
Expand your base of knowledge about AIDS/HIV with *AIDS Targeted Information/ATIN*. Every month, *ATIN* not only provides abstracts but also in-depth evaluations of the current published scientific literature on AIDS. Written by clinicians and researchers for clinicians and researchers, *ATIN* provides an authoritative command of the world's literature on AIDS. Published by Williams & Wilkins. Indexed. 12 issues per year. \$125.

To order call
1-800-638-6423

Sponsored by the American Foundation for AIDS Research (AmFAR).

Reader Service No 430

The rainforests — a view from above



SOME of the satellite images that provoke endless discussions about the true extent of deforestation in the Brazilian Amazon can be seen in *The Amazonian — A satellite's view*, a lavishly produced book with text in Portuguese, English and French by the Brazilian journalist, Liana John (published by Editorações Ltda., São Paulo; price 10,000 cruzeiros, approx. \$130). The great merit of the book is that it presents both satellite images and, facing them, photographs of the same region taken from the ground or air. This provides a striking illustration of the variety that characterizes the region, producing a portrait that is very different from the stereotype of the rain forest as a solid mass of green foliage.

The pair of pictures above show the Transamazon Highway (BR-230) in Pará. In the satellite picture (top) the highway crosses the Xingu River (dark blue), which flows for part of its course in a distinct zig-zag pattern as it sweeps over a series of rock faults. Branching out from the main road can be seen huge numbers of side roads, proliferating unknown to the government, and allowing settlers to destroy the forest without restraint. The accompanying ground picture shows a part of the highway as it was being built, illustrating how quickly forest alongside a new road is cut down.

If the book has a fault, it may lie in a certain defensiveness concerning problems such as this in the Amazon. Although pictures illustrating the flooding of vast areas by new hydroelectric schemes, the burning of forests and the pollution of rivers by mercury are presented, critical comment on the destruction is passed over with the remark that "it serves as a lesson in what not to do". But such defensiveness may be understandable in a multi-lingual book. Every Brazilian abroad has to deal with the angry question, "Why are you burning the forest?", even if that person has never set foot in the Amazon.

Ricardo Bonalume

Ricardo Bonalume is a science reporter for the *Folha de São Paulo* and *Nature's* correspondent in Brazil.

Selecting the answer

Mark Ridley

Adaptation and Environment. By Robert N. Brandon. Princeton University Press: 1990. Pp.214. \$29.95.

WHAT is the unit of selection? This may look like a single question, but it is actually two. In one sense, the answer is genes. Copies of genes can last almost for ever, and natural selection can adjust gene frequencies directly between generations in a way that it cannot adjust organism, genome or genotype frequencies. The causal priorities are uncomplicated: genes copy themselves, organisms do not, and 'organism selection' cannot work. You cannot adjust the frequency of something that dies every generation.

The second issue is as follows. For many kinds of social behaviour, we can ask whether individuals or groups benefit. Territoriality, for example, may exist because groups of territorial birds do not over-eat their food supply, or because one territorial bird gains over another by monopoly power. Either way, the behaviour will have evolved because selection favours genes for territoriality. The issue is whether those genes have been selected via an advantage at a group or an individual level.

We can work out the level of advantage by an ecological study of selection pressures. Roughly speaking, we measure the fitness of territorial and non-territorial individuals, and of groups, and then inspect the numbers. If non-territorial individuals out-reproduce territorial individuals, whereas territorial groups out-reproduce non-territorial groups, then we can attribute territoriality to group selection. Brandon has considered more carefully what has to be done. The comparisons have to be in a common environment (for otherwise the higher fitness of non-territorial individuals might just be because they are living in a better place), and Brandon also develops the exact concept of environment that is needed for the argument to work.

So two accounts of natural selection are possible. We can measure the fitness of different kinds of organism (and group, and so on) and see which kind selection favours, or we can divide the organisms (and so on) up according to the genes they possess and see which gene is reproducing faster. On the analysis I accept, these accounts are equivalent: one is not superior to the other.

Brandon argues for a different position. He is (I believe) mainly right about organisms and groups, but fundamentally unsound on genes. He argues that the

hierarchy of organisms, families and groups is the correct way to think about levels of selection, and that gene selectionism is inferior — "it has no explanatory value", and provides only "post hoc redescription of selection scenarios". He arrives at his conclusion by arguing that a full ecological study of the selection pressures on organisms, groups and so on will reveal what level an adaptation benefits: whereas just saying an adaptation benefits the genes does not reveal whether it also benefits organisms, groups or what.

But Brandon is comparing a complete ecological analysis for organisms and groups with only the abstract elements of gene selectionism. A full study of why one kind of gene is favoured over another should reveal the nature of the advantage in equal detail. Gene selectionists, confronted by a territorial species, do not just remark "Ah, the genes for territoriality are being favoured", before dapperly gliding on to another labour — any more than Brandon's ecologists observe (languidly, from an armchair) that it is because territorial organisms (or groups) are producing more offspring.

I therefore do not accept Brandon's case against gene selection. Properly applied, and in equal depth, the two methods give the same answers. Sometimes, one will be more convenient than the other. Brandon often reasons from the example of heavy-metal tolerance in plants, and here the gene's-eye view may well have no advantage. But it does in other cases. Inclusive fitness, which Brandon does not consider, is a strong example. In social behaviour, it is easy to count up the number of copies of a gene and see whether it is favoured, whereas counting organism equivalents (some of me, suitably discounted, plus a quarter of Aunt Agatha, also discounted, plus . . .) is enough of a puzzle to have led several teams of professional biologists into notorious absurdities.

Brandon, too, would have benefited from the gene's-eye view in one of his



This sixteenth-century village in Germany is reproduced from *River Pollution: An Ecological Perspective* by S M Haslam and shows how small settlements could often be hygienically designed with regard to pollution. Drinking water is clean, from the well on the left; pollution from hanging sheepskins takes place downstream from fishing (shown by the poles for nets). This book analyses the biological impact of pollution and discusses ways of monitoring damage. Published by Belhaven Press, price is £47. □

chapters. He discusses the evolution of "weak altruism", in which an individual pays a cost while giving a benefit to all members of its group, including itself. The benefit to each recipient exceeds the cost, so the individual weak altruist produces more offspring than if it were selfish. The problem is whether group selection is needed for weak altruism to evolve. The crucial test is the "mutation test". Imagine the mutation of a selfish into a weakly altruistic individual. If the mutant's relative fitness goes up, the trait will evolve by individual advantage. As any well-educated biologist knows, the mutation test (correctly applied) gives the right answer.

But Brandon does not like this test. He finds some conditions under which weak altruism in one group is favoured (according to the mutation test) if there are lots of other groups in the population, but not if there are only a few other groups. Brandon calls this "a strange result", and, moving into italics, says that, according to the mutation test, "whether or not [weak altruism] is individually advantageous depends on the presence or absence of other groups *between which there are no fitness-affecting interactions*". But, alas, there are. The strange result is nothing stranger than frequency-dependent selection; and the "fitness-affecting interaction" arose when Brandon calculated

the mean fitness. Suppose we want to calculate the mean fitness of the population in the United States. Suppose also there is a new strain of superfecund chaps in Miami, who reproduce 1,000 times per minute. We may have no interaction with them; we may never see them or even know of their existence. But our relative fitnesses are influenced by them.

Adaptation and Environment is a philosophical study. Brandon aims to give an accurate account of natural selection and the concept of environment needed by that theory. He tackles the issue of tautology and the nature of adaptationist explanation, as well as the unit of selection. The book is written for philosophers, and professionals will need to know what is in it. Evolutionary biologists who enjoy these questions should find it, like the assistant curate's egg, stimulating in parts. □

Mark Ridley is in the Departments of Anthropology and Biology, Emory University, Atlanta, Georgia 30322, USA.



A hidden bouquet — These images of an iris (above) and columbine (below) were made by placing both flowers between an X-ray source and film. Originally published in the *Smithsonian* magazine, they are now reproduced in *Editor's Choice: Smithsonian*, a book celebrating the magazine's 20th birthday. An assortment of articles covering science, history and the arts provides a 'best of' collection. Published by Smithsonian Institution Press, price is \$35. □

Astrophysical Sun

D. O. Gough

Solar Astrophysics. By Peter V. Foukal. Wiley Interscience: 1990. Pp.475. £43.65, \$62.60.

SOLAR physicists are frustrated people. The reason is that so much more is known about the Sun than about any other star. Astrophysicists are free to make quite simple stellar models, and can compare their theoretical predictions with the relatively few data that are available. At least a modicum of success is inevitable, and satisfaction is therefore guaranteed. By contrast, for almost every theory of a solar phenomenon, there is an observation to contradict it.

So numerous are the solar data that many a solar physicist has been feared lost in the morass of detail. Peter Foukal's guide to solar physics is aimed at preventing that loss from continuing, partly by informing scientists in related fields, but perhaps mainly by introducing the subject in a gentle manner to both undergraduate and graduate students. His book provides a refreshing survey of both current and mature thinking on what is known at present about our star.

As the title suggests, Foukal has written the book with astrophysicists in mind. The many areas of basic physics pertaining to solar phenomena are simply described, the most elementary from first principles, although the occasional simplification beyond the bounds of truth is not to my taste. Equations are presented that permit one to make order-of-magnitude assessments of conditions in the Sun and its atmosphere. Only rarely is one taught how to use them more precisely; there is not enough space between two covers to deal more quantitatively in a useful way with all the issues that are discussed.

But to my mind what is most useful are Foukal's discussions in coherent physical terms of the many observations of the Sun. Inevitably, there must be a risk of over-indulgence, but the author successfully steers a careful course between excessive detail and over-generalization. It is evident that he has invested a great deal of effort in organizing the material to make it palatable, and often even digestible. Foukal is an authority in his subject, and one can rest assured that the explanations given, when they are not his, are those that are accepted by most scientists working in the field. For many readers, however, Foukal's own deliberations would probably be more interesting.

As is pointed out in the preface, there are several good monographs and review articles on those areas of solar research where efforts are currently most concentrated. But these are usually rather



Norman Rockwell's painting of an astronaut's first step onto lunar soil — based on Neil Armstrong's in 1969 — is one of a collection of artists' impressions of the American Space Program published in *NASA: Visions of Space* by Robin Kerrod. Produced in colour, the paintings, which are all from the NASA Art Program, range from the technically precise to the abstract. Publisher is Prion, price is £17.95. □

specialized, and therefore they do not address the most obvious grand questions. *Solar Astrophysics* is intended to alleviate that deficiency by offering a more general discussion. Thus, for example, one can read in simple terms about how on a timescale of days solar irradiance declines when a sunspot comes into view, yet on a timescale of years the irradiance is greatest at times of maximum sunspot coverage. A convincing explanation is not given, but this is an indication of the limitations of our understanding, rather than of a deficiency in the author's ability to summarize the literature.

There is no doubt that *Solar Astrophysics* is an up-to-date, authoritative survey of the subject. It is well written in an easy style that is readily accessible to students, providing an excellent introduction to those who have not yet suffered from the welter of information that they might subsequently encounter. Moreover, it also provides a useful work of reference for those on the periphery of the subject, and perhaps also for some inside it, though rarely, if ever, does it refer to original sources of information. I would certainly recommend it as background reading to any student pursuing a serious course in solar physics. □

D. O. Gough is at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

Progress towards a quantitative understanding of Antarctic ozone depletion

Susan Solomon

The possibility that the stratospheric ozone layer could be depleted by half at certain latitudes and seasons would have been deemed a preposterous and alarmist suggestion in the early 1980s. A decade later, the statement is acknowledged as proved beyond reasonable scientific doubt. Observations of the composition of the Antarctic stratosphere have established that the chemistry of this region is highly unusual because of its extreme cold temperatures, leading to a greatly enhanced susceptibility to chlorine-catalysed ozone depletion.

OZONE has a critical role in the Earth's ecological balance owing to its strong absorption of biologically damaging incoming ultraviolet light¹. It has long been recognized that chlorine can destroy stratospheric ozone^{2,3}, and that the use of chlorofluorocarbons has led to a pronounced increase in the stratospheric chlorine content. Since the late 1970s, several chlorofluorocarbons have been measured at remote sites around the world. The observed rates of systematic increase of compounds such as CF₂Cl₂ (CFC-12) and CFCl₃ (CFC-11) in the atmosphere have been shown to be consistent with their known anthropogenic production rates worldwide⁴, verifying that human activities represent the only significant source of these molecules and that their destruction occurs uniquely in the stratosphere. By the early 1980s the chlorine content of the Earth's stratosphere had risen to about 2.5 parts per 10⁹ by volume (p.p.b.v.) compared to the natural level of about 0.6 p.p.b.v., due largely to the oceanic release of CH₃Cl⁵.

The discovery of the Antarctic ozone hole in 1985⁶ was met with a certain amount of arguably justifiable scepticism, motivated in part by the belief that the lower Antarctic stratosphere would not be the location where ozone depletion would first be manifested. Rather, it was thought that ozone near 40 km in mid-latitudes would show the first signs of change. Consider the fate of a chlorine atom released by stratospheric photolysis of CF₂Cl₂



The atomic chlorine thus produced can participate in a catalytic cycle that destroys ozone very effectively near 35–45 km



But because most of the total stratospheric ozone content resides at much lower altitudes (the maximum in ozone densities typically lies near 15–25 km), this and other ozone-destroying chemical cycles involving chlorine were predicted to lead to column-integrated ozone changes of perhaps 5% sometime near the middle of the twenty-first century using models considering the detailed gas-phase photochemistry of the atmosphere⁷. Such theoretical predictions stood in stark contrast to the remarkable 50% column change reported for the Antarctic in 1985 by Farman and co-workers⁵. Indeed, it was thought that the chemical lifetime of the bulk of the ozone layer above Antarctica was of the order of many years, and thus should show little sensitivity to chlorine increases. Nevertheless, Farman *et al.*⁶ suggested that chlorine might be the cause of the ozone depletion they observed, and noted that the ozone decrease found above their station at Halley Bay displayed a temporal correlation with the increasing burden of atmospheric chlorine. Figure 1 is adapted from their results, and shows the observed decline in the October monthly mean ozone column above Halley Bay over the past three

decades or so. Also shown are data from the South Pole station^{8,9}. These observations along with other data¹⁰ (summarized in detail in a recent international scientific review⁵) verified the remarkable change in Antarctic ozone, now widely referred to as the ozone 'hole'. Figure 1 also shows the temporal evolution of the total chlorine content of the troposphere and stratosphere, calculated from their release rates and observations of natural and man-made compounds^{4,5}. It is reasonable to expect some lag between the chlorine content of the troposphere and that of the polar lower stratosphere because of the time taken to transport chlorine from the troposphere to the polar stratosphere. Recent studies (A. F. Tuck, personal communication) suggest that the transport time is between three and five years, based on observations of carbon dioxide in the Arctic stratosphere. It is interesting to note that the ozone depletion became apparent between 1975 and 1980, a time when the stratospheric chlorine content was increasing particularly rapidly (see Fig. 1).

The ozone hole is a seasonal phenomenon that is tied to the dynamic meteorology of the Antarctic atmosphere. During winter and early spring, the Antarctic has the coldest temperatures found in the Earth's stratosphere, some 10–15 K colder than

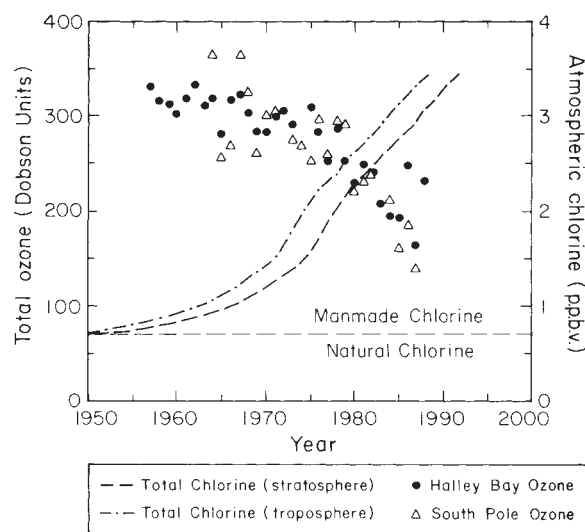


FIG. 1 Total ozone observations above Halley Bay⁶ and the South Pole^{8,9}, showing the development of the ozone hole. Total chlorine abundances in the troposphere and stratosphere are also indicated, based on an assumed 3.5-year lag time between the two atmospheric regimes (ref. 5 and A. F. Tuck, personal communication). Natural chlorine abundances are believed to be ~0.6 p.p.b.v., with man-made chlorine making up the remainder of the stratospheric chlorine budget⁶.

those of the Arctic in the same seasons. This interhemispheric temperature difference is believed to be due largely to differences in the surface topography of the two hemispheres (in particular, to the lack of surface topographic features poleward of 50°S and equatorward of the Antarctic continent). These lead to fewer large-scale wave disturbances at sub-polar latitudes and hence to less wave-induced mixing of polar air with the warmer air prevailing at mid-latitudes for the Southern Hemisphere. These conditions, in turn, result in a strong temperature contrast between the Antarctic and the mid-latitudes of the Southern Hemisphere. It has long been known that the steep temperature gradient is accompanied by rapid circumpolar flow and a relatively isolated polar vortex¹¹. As will be discussed later, the degree of that isolation is a critical issue for quantitative studies of the budget of Antarctic ozone depletion as well as its implications for lower latitudes.

Observations at several Antarctic sites^{5,9,12} established that the ozone decline occurs largely during the months of September. Figure 2 shows that the ozone abundances observed presently in early September approach historical levels, but that the ozone decreases rapidly during September, reaching a minimum in early October. Ozone abundances found currently in October are about 50% lower than the historical values of about 300 Dobson Units⁵. This seasonal 'hole' is then filled in to a substantial degree by the rapid influx of ozone-rich air from lower latitudes that accompanies the dynamical breakdown of the winter polar stratospheric vortex^{5,11,13}, typically in October or November.

Although suggestive, the correlation depicted in Fig. 1 demands much stronger proof, which can only be provided by observations of the chemistry taking place in the Antarctic polar vortex. But when the ozone hole was discovered, such data were sorely lacking.

Theories

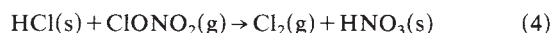
In the absence of observations, several distinct, and in some ways quite opposing, theories were advanced. The leading candidates among these became known as the solar-cycle theory¹⁴, the dynamical theory^{15,16} and the chlorofluorocarbon theory^{17,18}. More detailed descriptions and comparison of the competing theories have been presented elsewhere^{5,19}.

In brief, the basis for the solar-cycle theory was the notion that reactive nitrogen compounds (NO and NO₂), like chlorine, can destroy stratospheric ozone quite effectively, and that reac-

tive nitrogen is produced in abundance in the mesosphere and lower thermosphere (that is, above ~60–70 km) during periods of high solar activity. The nitrogen compounds are rapidly destroyed in the sunlit mesosphere, but might be transported all the way to the lower stratosphere in the dark polar winter. An unusually active phase of the 11-year solar cycle could then be suspected of enhancing the stratospheric abundances of reactive nitrogen compounds in the polar vortex, which could destroy ozone photochemically following the return of sunlight to the polar cap in September. This theory was effectively eliminated by observations showing that the abundances of nitrogen oxides were exceedingly low, not high, in the Antarctic vortex, as compared to lower latitudes or Arctic regions^{20–22}. Indeed, such observations actually supported the chlorofluorocarbon theory, as discussed below.

The leading dynamical theory rested on the possibility that solar heating taking place in an unusually cold environment such as the Antarctic lower stratosphere could lead to net radiative warming. Following the simple but largely correct notion that heated air rises, such heating could bring ozone-poor air upward from the troposphere, decreasing the local ozone column abundance. This theory was, however, found to be inconsistent with both the detailed vertical structure of the observed depletion⁵ and observations of background aerosols and other tracers of atmospheric flow patterns, such as N₂O, in the polar vortex^{23,24}. These data established that large-scale motion could explain neither the magnitude nor the vertical extent of the September ozone depletion. Indeed, the tracer observations suggest that the meridional flow pattern in the Antarctic spring is largely characterized by downward rather than upward motion.

The chlorofluorocarbon theory began with the recognition that important chemical reactions might be taking place on cloud surfaces in Antarctica, changing a long-standing emphasis on gas-phase stratospheric chemistry to include perturbations by heterogeneous processes. It has long been recognized that the extreme cold temperatures of the Antarctic stratosphere lead to the formation of polar stratospheric clouds there, in much greater frequency and thickness than found elsewhere on Earth, even in the Arctic²⁵, but these clouds were considered little more than a scientific curiosity for many years. Following the discovery of the ozone hole it was suggested¹⁷ that heterogeneous reactions such as



could convert chlorine from the rather inactive compounds, HCl and ClONO₂ into a far more reactive species, Cl₂ (s, solid; g, gas). This molecule will in turn readily dissociate in the presence of sunlight to form atomic chlorine which can then go on to destroy ozone during the sunlit period in September. Heterogeneous reactions would be expected to take place in the coldest part of the Antarctic stratosphere, where the clouds form (at ~10–25 km altitude). Such processes have the important effect of altering the altitude range over which chlorine compounds are capable of ozone destruction from the ozone-poor altitudes near 35 to 45 km to lower, ozone-rich altitudes, thus yielding far larger total column ozone depletion than anticipated based on standard gas-phase photochemistry alone. Laboratory studies have demonstrated^{26–28} that reaction (4) proceeds rapidly on ice surfaces but not in the gas phase. Other heterogeneous reactions such as



are also believed to yield important perturbations to the gas-phase photochemistry.

It has been recognized that the polar stratospheric cloud themselves are probably composed in part of nitric acid

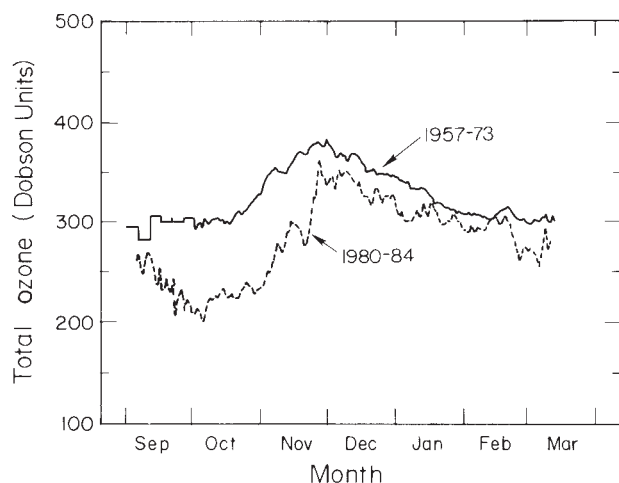


FIG. 2 Comparison between the seasonal cycle of total ozone historically observed at Halley Bay and that of the early 1980s^{5,12}. A rapid decline occurs during September and persists into October and November, with smaller changes in other seasons. In October 1987 the average total ozone was even lower, only 163 Dobson Units⁵.

hydrates²⁹⁻³¹, particularly the trihydrate ($\text{HNO}_3 \cdot 3\text{H}_2\text{O}$)^{29,30}. Laboratory studies have confirmed that nitric acid trihydrate will form a solid at temperatures considerably warmer than water ice, thus enhancing the frequency of occurrence as well as the vertical and latitudinal extent of the reactive surfaces in the stratosphere. Also, reactions such as (4)–(7) and the formation of nitric acid trihydrate clouds affect the nitrogen chemistry of the stratosphere as well as the chlorine chemistry. In particular, reactions (4) and (5) convert inactive chlorine to very active forms (hereafter referred to as ClO_x), while at the same time converting reactive nitrogen (in the form of the NO_2 contained in ClONO_2) to HNO_3 , an exceedingly inert form of nitrogen in the winter lower stratosphere. Reactive nitrogen (NO_x) plays an essential role in chlorine chemistry due to the chemical reaction



(where M is any third body), which is the primary process by which active chlorine can be converted back to an inactive form. The timescale for this conversion is limited by the amount of available NO_2 . The abundance of NO_2 thus controls the length of time during which the chlorine activated by surface reactions remains available to destroy ozone and thereby plays a critical part in determining the extent and duration of ozone depletion. The formation of HNO_3 by reactions (4)–(7) has the net effect of reducing the abundances of the more reactive forms of nitrogen (including NO_2). Further, the physical process of formation of the nitric acid trihydrate cloud particles represents an even more effective means of reducing reactive nitrogen, because this removes it from the gas phase. Finally, the possibility that cloud particles containing HNO_3 may attain sizes sufficient for sedimentation²⁹ implies that reactive nitrogen may be removed from the stratosphere altogether, which further enhances the effectiveness of chlorine chemistry for ozone destruction. Observations have indeed shown evidence for denitrification and dehydration of the Antarctic stratosphere^{32,33} apparently through sedimentation of cloud particles, although considerable debate remains concerning the detailed microphysical mechanism for cloud-particle growth^{34,35}. This issue may be of particular importance for studies of Arctic ozone depletion (see, for example, ref. 34).

These heterogeneous processes thus act on both the chlorine and nitrogen chemistry of the Antarctic lower stratosphere and

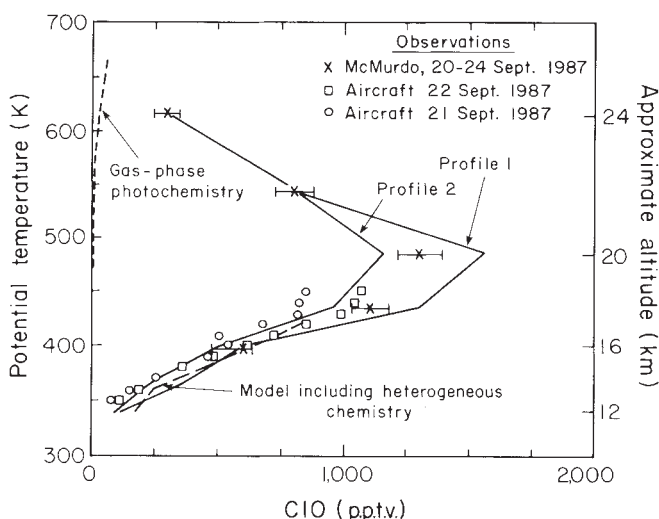
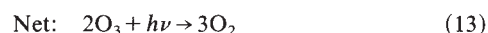
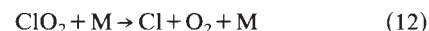
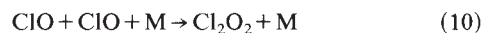
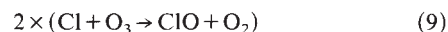


FIG. 3 Observations of the ClO profiles above McMurdo Station⁴⁷ and near the Palmer Peninsula⁴⁰ in September 1987. Model calculations considering only gas-phase photochemistry⁴⁸ and one including a parameterized treatment of heterogeneous chemistry⁴⁹ are shown for comparison. Profiles 1 and 2 are fits to the measured ClO distribution discussed in the text.

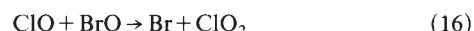
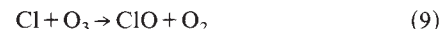
set the stage for remarkable enhancements in the abundances of ozone-damaging forms of atmospheric chlorine, at the expense of relatively benign forms. The following catalytic cycles are believed to be responsible for the ozone loss



Under cold sunlit conditions, the rate limiting step in this catalytic cycle³⁶ involving reactive chlorine and sunlight is the reaction of ClO with itself to form the ClO dimer, Cl_2O_2 . The rate of chemical ozone loss owing to this cycle can then be written as

$$d(\text{O}_3)/dt = -2k_{10}(\text{ClO})(\text{ClO})(\text{M}) \quad (14)$$

Reactions involving the coupling between chlorine and bromine¹⁸ also contribute to the ozone destruction rate, particularly through the reaction cycle



Present observations^{5,37,38} suggest that the abundance of atmospheric bromine is about 10 parts per 10^{12} by volume (p.p.t.v.), and that much of the total bromine in the Antarctic vortex exists in the form of BrO. Although the contemporary burden of stratospheric bromine is believed to be largely natural (because of surface releases of CH_3Br), anthropogenic emissions of bromine-containing halons are rapidly increasing⁵. Enhanced ClO abundances increase the rate of ozone loss because of the coupling between chlorine and bromine¹⁸, as the rate limiting step in this cycle is the reaction of ClO with BrO. The rate of local ozone loss owing to both cycles is

$$d(\text{O}_3)/dt = -2\{k_{10}(\text{ClO})(\text{ClO})(\text{M}) + k_{16}(\text{ClO})(\text{BrO})\} \quad (18)$$

and is defined, in principle, by the abundances of ClO and BrO along with laboratory measurements of the reaction rate constants³⁶⁻⁴². The observed rate of ozone decrease near 20 km in the Antarctic vortex is about 2% per day during September⁴³⁻⁴⁵. An order of magnitude estimate adopting present measurements of the relevant reaction rate constants and using equation (18) suggests that such a loss rate is expected for lower stratospheric ClO abundances of the order of 1 p.p.b.v. This number is about one hundred times greater than predictions by gas-phase photochemical models, indicating the massive perturbation to the photochemistry required if chlorine chemistry is to be the key factor in the existence of the ozone hole. Such a perturbation should be readily apparent not only in ClO observations, but also in measurements of related compounds involved in the chemistry such as OClO, HCl, ClONO_2 and NO. Extensive field campaigns in the Antarctic were carried out in 1986 and 1987 to obtain the needed data.

Towards a qualitative understanding

Most of the measurements of Antarctic chemical composition have been obtained either from aircraft observations carried out on flights from Chile to Antarctica⁴⁶ or from ground-based studies at McMurdo Station (78° S, 168° E).

Equation (18) shows that the critical chemical species involved in ozone destruction is likely to be ClO, a reactive radical that has now been measured in Antarctica by two completely independent methods: ground-based microwave emission^{39,47}

and a resonance-fluorescence technique conducted from aircraft⁴⁰. Figure 3 shows the vertical profiles of ClO obtained from the two methods compared to a two-dimensional model prediction involving only gas-phase chemistry⁴⁸, and a trajectory model in which heterogeneous chemistry is simulated by considering reaction probabilities measured in the laboratory for the reactions indicated above, along with observed local temperatures to infer the presence or absence of clouds⁴⁹. The two sets of measurements taken together demonstrate that the chemistry of the Antarctic stratosphere is remarkably perturbed, with maximum ClO abundances of the order of 1 p.p.b.v., qualitatively consistent with that required to explain the observed rate of ozone loss as indicated earlier. The model simulation neglecting heterogeneous chemistry is unable to reproduce the observed ClO profile, whereas the model including heterogeneous reactions is consistent with the measured abundance, and perhaps more importantly, with its vertical structure (although it must be noted that the calculated ClO in this case is dependent on concurrent reactive nitrogen measurements and estimates of the rates of heterogeneous reactions used to constrain the model). It is interesting that the ClO observations from McMurdo Station suggest maximum values somewhat greater than the mean of those found near the southernmost portion of the aircraft flights (~72° S), even when likely adjustments to the aircraft measurement calibration are considered⁵⁰. This discrepancy, if real, suggests potentially important spatial gradients in the ClO abundance in the polar vortex. This point will be discussed in more detail below. Profiles 1 and 2 represent two possible fits to the observed ClO profiles that will also be further discussed below.

A closely related molecule to ClO is OClO, formed through one of the reaction channels of ClO with BrO^{18,51}. Both ground-based⁵² and airborne⁵³ remote-sensing measurements, using near-ultraviolet and visible absorption spectroscopy, have shown that OClO is enhanced above Antarctica during September by about a factor of one hundred compared to other latitudes or model predictions, providing another independent measure of the highly perturbed state of chlorine chemistry in the Antarctic stratosphere.

Measurements of the column abundances of HCl and ClONO₂ are available from both ground-based and airborne studies using high-resolution infrared absorption methods^{54–56}. These reveal greatly diminished values in the polar vortex as compared to lower latitudes, supporting the notion that the elevated abundances of ClO there are due to conversion from these less reactive compounds. Further, measurements of the column abundances of NO₂ and HNO₃^{20,21,53,54,56} as well as *in situ* observations of NO₂²² and total reactive nitrogen abundances³² near 20 km all demonstrate that the reactive nitrogen species are greatly suppressed in the Antarctic stratosphere, relative to other latitudes, consistent with the perturbations expected from cloud chemistry.

Taken together, this suite of independent measurements of numerous chemical compounds confirms the important role of heterogeneous chemistry in the Antarctic stratosphere beyond reasonable scientific doubt. The measured abundances of ClO are qualitatively consistent with those required to explain the rate of Antarctic ozone loss obtained in September 1987.

The quest to be more quantitative

It is tempting to approach quantitative evaluation of the ozone loss as a basic problem in chemical kinetics^{41,47,57,58}. In a well controlled laboratory environment, the measured rate of ozone loss (left-hand side of equation (18)) could be readily compared for consistency to that of known photochemical mechanisms, that is, the measured reaction rate constants and ClO and BrO abundances (right-hand side of equation (18), hereafter referred to as the photochemical loss rate). Some analyses using Antarctic ClO and BrO measurements, along with present estimates for reaction rate constants suggest discrepancies as large as 35%^{57,58},

and have raised concerns regarding possible additional chemical ozone destruction mechanisms or other gaps in the understanding of the ozone depletion process. In the following, such a kinetic analysis is attempted and numerical and conceptual problems that limit its interpretation described.

The abundance of ClO is expected to display a pronounced diurnal cycle, with a maximum near noon, due largely to the photolysis of Cl₂O₂; such a diurnal change is apparent in the ground-based measurements from McMurdo Station³⁹. Thus the time-dependent diurnal cycle of ClO must be considered when evaluating the ozone loss throughout the day using equation (18). *In situ* ClO (and BrO) measurements are made at only two local times (northbound and southbound legs of the flights). These point measurements must be combined with model calculations to infer abundances for other local times. Also, the solar illumination generally varies strongly with latitude in the polar cap region during late winter and spring, with lower latitudes receiving more solar radiation than those close to the pole. Assuming that the atmosphere is either stagnant or that flow is restricted to perfect latitude circles so that the diurnal photochemistry repeats exactly from one day to another, the photochemical rate of ozone loss can be evaluated and compared to concurrent ozone observations taken during the month of September. Clearly, the modelled diurnal cycle of ClO and the assumed symmetric flow pattern represent possible sources of uncertainty; this will be discussed further below.

Figure 4 shows measured and photochemical profiles of ozone-loss rates in mid to late September 1987. One set of 'measured' ozone-loss rates was deduced from a detailed study of the temporal evolution of chemical tracer fields in the Antarctic vortex⁴⁴ (discussed further below), and another was inferred from a simple linear trend analysis of balloonsonde data above McMurdo Station over the period from 15 September to 28 September (D. J. Hofmann, personal communication). The two data sets both show that the maximum ozone change occurs near 17 km, dropping rapidly to zero above about 22 and below 12 km. The two sets of ozone data display a qualitative resemblance to the measured noon ClO profiles of Fig. 3 and to one another, but there are important differences, suggesting that even the rates of measured ozone decline may be difficult to characterize. Also shown in Fig. 4 are calculations of the photo-

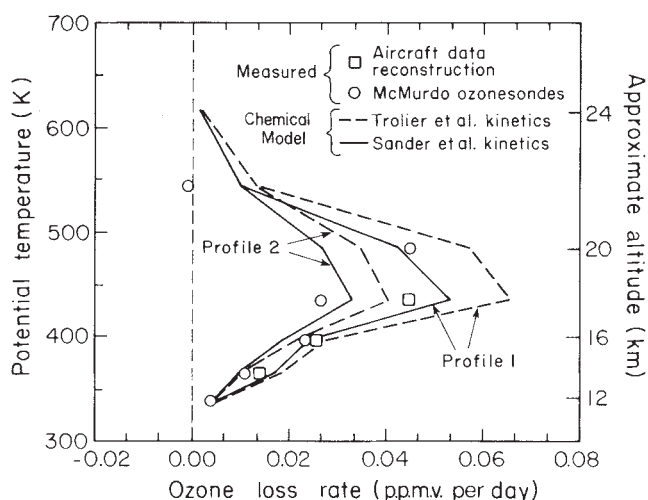


FIG. 4 Profiles of 'measured' and 'photochemical' ozone-loss rates above Antarctica in September 1987. Measured ozone-loss rates from a reconstruction analysis based on long-lived tracer data are shown⁴⁴ along with those obtained from a linear trend of the ozonesonde measurements above McMurdo Station (D. J. Hofmann, personal communication). Photochemical ozone-loss rates from model calculations including detailed diurnal chemistry^{61,62} and using the noon ClO vertical profiles shown in Fig. 3 and different kinetic rate constants^{41,42} are also shown for comparison.

chemical ozone-loss rate around 20 September at 75° S based on a simple photochemical model^{59,60} constrained to yield one of the two noon ClO profiles indicated in Fig. 3, and including no reactive nitrogen chemistry for simplicity. Reactions involving the coupling between chlorine and bromine were considered, and the assumed reactive bromine abundance was 7 p.p.t.v.⁴⁰. Profile 2 of Fig. 3 agrees best with the published aircraft ClO measurements⁴⁰ but results in ozone-loss rates below those observed (particularly when the slower kinetic rate for reaction (10) is adopted⁴¹). On the other hand, calculations using Profile 1 of Fig. 3, or the faster kinetic rate, seem reasonably consistent with the observed rate of ozone decrease, at least in mid-September, although some authors have suggested larger discrepancies earlier in the spring season⁴¹. This and other calculations^{41,47,49,57} suggest that chlorine-bromine chemistry accounts for about 20% of the calculated ozone decline, and the chemistry of the ClO dimer makes up the remainder of the photochemical ozone destruction rate. Changes in the calibration⁵⁰ of the airborne measurements support ClO abundances substantially larger than indicated by Profile 2 and perhaps alleviate some concerns regarding such simple ozone budget studies. In spite of this tempting improvement in agreement between the left- and right-hand sides of equation (18), it is clear from Fig. 4 that the uncertainties in kinetic rate constants and ClO abundances remain large. Further, the conceptual limitations inherent in such a comparison of the photochemical loss rate and the inferred measured ozone-loss rate must also be considered. Figure 5 is a schematic representation of some of the other complications that will now be discussed.

The Antarctic stratosphere is far from stationary. Rather, the winds are predominantly from the west to the east around the Antarctic winter polar vortex and are of the order of tens of metres per second, so that air parcels revolve around the continent in approximately one week. Thus ozone loss occurs not just in the few restricted locations where the ClO observations are currently available, but rather in moving air parcels that may be exposed to a range of photochemical conditions. The implications of the changing transport and photochemistry experienced by moving air parcels⁴⁹ must be considered when making simple photochemical calculations such as those shown in Fig. 4.

The motion of air parcels in the real stratosphere is not generally restricted to perfect latitude circles. Rather, atmospheric waves disturb the polar vortex, distorting the flow field and causing substantial north-south excursions. The daytime abundance of ClO is critically dependent on the photolysis of the ClO dimer. Near noon at 78° S in mid-September, ~25–30% of the active chlorine is expected to reside in the form of the dimer near 20 km, with most of the remainder in the form of ClO. At lower latitudes near 65° S, increases in solar illumination are expected to result in smaller dimer concentrations and larger ClO abundances, so that an air parcel moving from 78° S to 65° S near the 20-km level would be expected to display ~10% increase in ClO even if the total ClO_x is conserved following the motion. Model studies including detailed treatments of stratospheric flow based on available meteorological data indicate that such motions are likely to be quite important in determining the solar illumination experienced by moving air parcels, and by extension, the ClO variability and the photochemical ozone-loss rate in Antarctica^{49,61}. These important effects are illustrated by the differences between path A and path B as shown in Figure 5.

Perhaps of most importance is that the likelihood of air parcels encountering a polar stratospheric cloud seems to be distributed non-uniformly around the Antarctic vortex. Satellite observations of polar stratospheric clouds during September 1987 have revealed that they occur most frequently near the eastern side of the Antarctic Peninsula⁶² (for example in the area of the Weddell Sea) whereas the aircraft measurements of ClO were confined to the western side (largely for safety reasons). This

suggests that air parcels are likely to be exposed to polar stratospheric clouds shortly after observation by the aircraft. As illustrated schematically in path B of Fig. 5, such exposure is expected to lead to temporary enhancements in ClO_x and hence in ClO if there is any ClONO₂ present in the air parcel (through, for example, reaction (5)). Observed longitudinal variations in the frequency of occurrence of polar stratospheric cloud therefore suggest that ClO abundances near the Weddell Sea may well be larger than those obtained from the aircraft flights. Such possible spatial inhomogeneities pose an important obstacle to the use of spatially limited data to infer the ClO content in air parcels moving around the Antarctic continent.

Having discussed several conceptual barriers to accurate determination of the right-hand side of equation (18), it is useful to consider some of the difficulties associated with accurately defining the 'measured' rate of ozone loss based on measurements over a ground-based station or along a series of aircraft flight tracks (left-hand side of equation (18)). As in the discussion above, the task is quite simple if air parcels are stationary, but becomes rather difficult when the vortex is displaced or distorted (under dynamically active conditions). Ozone abundances in the Antarctic polar lower stratosphere generally decrease with increasing latitude, so that horizontal displacements of the vortex relative to a particular point can cause changes in the local ozone abundance that do not necessarily reflect chemical loss. Horizontal motion of the vortex thus makes it difficult to define a coordinate system from which to characterize the ozone-loss rate; certainly latitude alone is quite inadequate, as the vortex may be displaced many degrees of latitude in a single day⁶³.

These difficulties can be alleviated to some extent by using chemical and dynamical tracers such as N₂O and potential vorticity to define the coordinate system, rather than latitude⁴⁴. This approach makes use of a chemical compound or dynamically derived quantity that is roughly conserved following air parcel motions, thus providing the needed tracer or 'tag' to identify air parcels that can usefully be compared. But, a limitation of this approach is the extent to which the tracers used are indeed conserved. For example, small-scale mixing can influence the N₂O abundance and would imply that this and other species (such as ozone) may not be absolutely conserved for air parcels resolvable by the spatial grid used; similarly, radiative heating or cooling can change the value of dynamical tracers such as

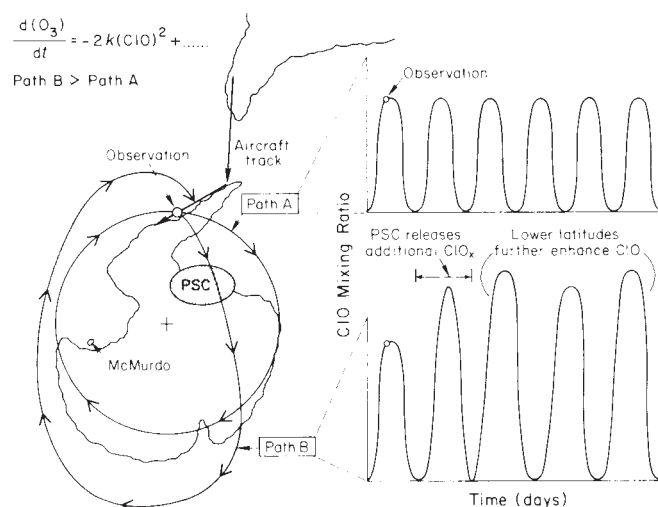


FIG. 5 Schematic diagram of possible air parcel paths around the Antarctic polar vortex and their resulting ClO content. Note that the rate of photochemical ozone loss is related to the square of the ClO abundance (equation (18)). Localized polar stratospheric clouds and latitudinal excursions of moving air parcels can have important roles in determining the ClO abundance and hence the photochemical loss rate.

potential vorticity or potential temperature over timescales longer than about ten days.

Large-scale vertical motion in the polar vortex must also be considered. Because ozone and total reactive chlorine (that portion of the total chlorine not in the form of chlorofluorocarbons) both have steeply increasing vertical gradients in polar regions, downward motion could represent a source for ozone at a particular altitude (by bringing down ozone-rich air from above). Such transport implies that measured local ozone trends cannot be equated to local chemical loss rates, particularly in dynamically active years. Further, downward motion may also enhance the photochemical loss rate by importing more reactive chlorine and/or HNO_3 and H_2O needed for the formation of clouds⁶³. There is, however, much debate at present regarding the speed and thus the importance of these meteorological processes. One study⁶⁴ used a variety of long-lived tracers with both latitude-based and tracer-based coordinate systems to attempt to characterize the flow through the longitude sector near the Antarctic peninsula from the aircraft measurements, and concluded that flow through the polar vortex was likely to be quite slow. The inferred direction of motion was downward. Another study⁶⁵ presented a very different view, noting that the observed behaviour of the ratio of tracers, $\text{CFCl}_3/\text{N}_2\text{O}$, provided evidence for strong descent in the Antarctic vortex. Some support for this view comes from aircraft³³ and satellite observations⁶⁵ of water vapour that suggest long-range transport of dehydrated air from the polar vortex to lower latitudes. This study spawned the notion that rapid downward flow may allow the vortex to act as a processor, exporting high ClO , low NO and low ozone to lower latitudes^{62,63}. A third analysis⁶⁶ used a coordinate system based on the measured abundance of the reactive species, ClO , from the aircraft data to deduce vertical transport and concluded that downward flow takes place particularly rapidly near the vortex boundary. In addition to downward flow, horizontal transport must also be considered, and the implications for the thermal budget assessed; such an analysis is beyond the scope of this review. Here I note only that these somewhat conflicting views of dynamical transport have important implications not only for attempts to quantify both the right- and left-hand sides of equation (18), but also for the possible export of ozone-poor air from Antarctica to lower latitudes.

If horizontal exchange of air and/or mixing in the vortex is possible, then latitudinal variations in temperatures, cloud frequency and hence in ozone trends should be further considered. For example, it is possible that chemical processes may take place particularly rapidly over the South Pole during some portions of the spring season, and that subsequent transport of depleted ozone to lower latitudes, including McMurdo Station or the Antarctic peninsula, can influence ozone trends at those locations. Figure 6 presents measured ClO abundances from the aircraft flights⁴⁰ along with those from McMurdo Station⁴⁷, as well as aircraft measurements of the latitude distribution of the sum of the column abundances of $\text{HCl} + \text{ClONO}_2$ ⁵⁶. The latter measurements were made on a different aircraft with greater range and flights went almost to the pole. Concurrent measurements of the HF column above the same flight tracks were constant to within about 7%; this and other tracer measurements suggest that the observed decrease in the sum of $\text{HCl} + \text{ClONO}_2$ at polar latitudes was probably because of heterogeneous reactions rather than air-mass changes or vertical transport. The observed steep gradients in ClO and in $\text{HCl} + \text{ClONO}_2$ appear to shift by as much as 5° latitude from one day to the next, illustrating the importance of disturbances and motion of the polar vortex in determining the chemical composition at any point in space and time. Nevertheless, the data show that significant decreases in $\text{HCl} + \text{ClONO}_2$ are likely to occur for latitudes poleward of the southernmost edge of the airborne ClO data, suggesting corresponding increases in ClO_x in the vortex, consistent with the relatively high ClO values measured at McMurdo. Larger ClO abundances in turn imply enhance-

ments in ozone-loss rates deep in the vortex, at least during the latter part of September when the entire polar cap is illuminated by the Sun^{45,67}.

In summary, substantial uncertainties remain regarding ClO abundances and kinetic rate constants involved in Antarctic ozone destruction. Further, presently available aircraft and ground-based data are limited to specific locations and times, so that attempts to deduce photochemical ozone destruction rates based on current data involve multiple extrapolations regarding diurnal, temporal and spatial behaviour of key reactive species. Physical processes such as longitudinal or latitudinal variations in polar stratospheric cloud frequencies and north-south excursions of air parcels represent important shortcomings in such extrapolations (see Fig. 5). The mixing of air parcels, vertical and horizontal motions, and radiative processes are other obstacles to quantitative evaluation of both the right- and left-hand sides of equation (18), namely, the observed decay rate of ozone itself and the derived rate of the photochemical processes that are likely to be driving that decay. Although additional chemical loss processes are possible, present uncertainties do not lead to a compelling case that they are required. The similarity between the vertical profiles of the calculated photochemical ozone-loss rates and those derived from limited measurements (Fig. 4) supports the view that halocarbon chemistry is responsible for much and possibly all, of the Antarctic ozone decline obtained during 1987, but a quantitative budget analysis is not possible with the present observations. Further spatial and temporal coverage (for example, from satellites, or long-duration aircraft or balloon measurements) is needed for progress to be made on the important issue of balancing the Antarctic ozone budget.

Interannual and interhemispheric differences

It has been noted that 1987 was a year with an extremely pronounced Antarctic ozone 'hole', whereas that of 1988 was much less severe⁶⁸. Figure 1 shows that modest fluctuations in total ozone were also observed in the years before the development of the Antarctic ozone hole, probably because of changes in the meteorologically driven transport of ozone to Antarctica. It is clear that ozone depletion is most severe in the coldest years^{68,69}, consistent with the view that ozone decline is closely linked to the formation of polar stratospheric clouds. A more quantitative analysis is not possible without detailed measurements of composition from one year to the next.

The rapid onset of the ozone decline in the late 1970s and early 1980s should also be considered, with a view towards

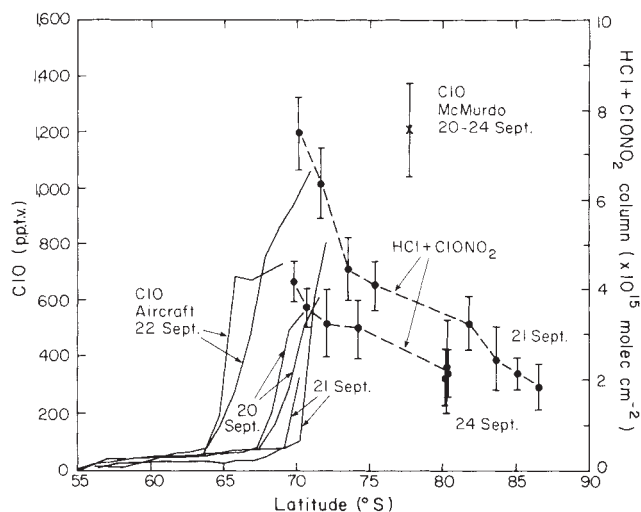


FIG. 6 Measurements of the latitude distribution of ClO near 18 km based on aircraft⁴⁰ and ground-based data⁴⁷. Observations of the sum of the column abundances of HCl and ClONO_2 ⁵⁶ are also shown.

determining whether the appearance of the ozone hole was due uniquely to the rapidly increasing atmospheric burden of man-made chlorine or whether some other atmospheric effect may also have contributed to triggering its development. Calculations of the ozone destruction based on known increases in chlorine content (Fig. 1) do display some nonlinearity in the expected ozone decline over the past decade, in part because the chemistry relating to the ClO dimer varies as the square of the ClO abundance (equation (18)). But simple calculations^{61,70} fall short of accounting for the decadal trend in Antarctic ozone when using only chemical processes and assuming that the extent and duration of cloud chemistry in previous years was identical to that of 1987. It seems that other factors may be involved. Because of the apparent relationship between ozone depletion observed in 1987 and heterogeneous chemistry, possible changes in temperature and hence in cloud frequency should be considered. Increases in tropospheric methane probably cause increases in stratospheric water vapour abundances⁷¹, but stratospheric cloud formation is generally expected to depend more strongly on temperature and the availability of HNO₃ than on water vapour concentrations. Detailed analyses of temperature data from Antarctic stations over about the past thirty years show no evidence for stratospheric temperature trends in September^{72,73}, suggesting that temperature changes did not trigger the onset of September polar stratospheric cloud occurrences and hence ozone depletion. Rather, it seems more plausible to consider that a positive feedback mechanism may be involved. Trends in temperature and in the frequency of cloud occurrence have been observed not in September, but in October^{74,75}. It is important to emphasize that these are apparently caused by the ozone hole rather than the reverse. As ozone is depleted, the primary source of heat in the stratosphere decreases, leading to colder temperatures⁷⁶ and increasing the occurrence of PSCs late in the season when normal ozone levels would have led to temperatures too warm for polar stratospheric clouds. Ozone depletion may thus extend the time during which the stratosphere remains cold enough for cloud formation, perhaps further enhancing the ozone decrease but not representing an independent causal mechanism. It has also been pointed out that discrete dynamical events (small-scale synoptic events) may cause significant cooling and thus play an important part in controlling the evolution of spring temperatures and photochemical ozone destruction in the vortex⁶³. More quantitative studies are needed to examine the interplay between these factors, but there is no observational support for any trigger for the ozone decline independent of the well known and precipitous rise in the amount of stratospheric chlorine.

Implications for other latitudes

Satellite observations of total ozone support the view that the Antarctic ozone hole does affect the ozone distribution over much of the Southern Hemisphere, at least through a dilution process association with the break-up of the vortex during the final stratospheric warming^{5,77}. The possibility that the vortex may process air and hence export ozone-poor air and photochemical perturbations to other latitudes has also been suggested, and it has been pointed out that a resulting transition or mixing zone may have a significant role in ozone depletion owing to its potentially large areal extent⁷⁸ compared to that of the interior of the vortex. Over the past ten years, satellite observations suggest that the ozone near 50°S latitude has declined by about 3–8%, depending on season⁵. It has also been suggested that the chemical reactions taking place on ice surfaces can occur on the liquid sulphuric acid surfaces present in the stratosphere at other latitudes^{79–81}, albeit much less rapidly. These considerations suggest that future ozone depletion at sub-polar latitudes will probably exceed theoretical predictions based on gas-phase chemistry alone, either through local heterogeneous processes and/or through transport from the polar vortex.

In the Northern Hemisphere, there is evidence for a 5–8% decrease in total ozone, with the largest depletions at the highest latitudes and during the winter and spring⁵. It is beyond the scope of this review to compare the two hemispheres in detail. There is abundant direct evidence for perturbed chemistry in the Arctic similar to that of Antarctica⁸², but the requisite cold temperatures are generally neither as widespread nor as long-lasting as in the spring season in Antarctica. This suggests that the conditions needed for ozone loss (namely, both cold temperatures and sunlight⁸³) are likely to be generally less effective in the Arctic than in Antarctica. For example, the minimum temperatures suggested by the National Meteorological Center stratospheric analyses⁸⁴ for the 30-mbar level (~24 km) were about –61 °C around 15 March in the Arctic during 1989 as compared to –78 °C for 15 September in Antarctica during 1988 (a relatively warm year); these values are fairly typical of each hemisphere. But fluctuations in temperature must also be noted. The Arctic spring seasons of 1967 and 1976 are the coldest in the historical record, which spans about three decades. In mid-March 1976 and mid-March 1967, observed minimum temperatures within the Arctic vortex at 30 mbar were only about –82 °C and –75 °C, respectively (ref. 84 and R. Nagatani, personal communication). These data show that the Arctic spring stratospheric temperatures can be comparable to those of Antarctica, at least at some altitudes. The chlorine content in 1967 and 1976 was too low for measurable ozone loss (Fig. 1), but it is reasonable to expect that significant ozone decreases in the Northern Hemisphere would be observed if such unusually cold conditions were reached in the present Arctic atmosphere.

Further research is needed to answer remaining quantitative questions regarding the detailed budget of Antarctic ozone loss and to address the challenging problems of mid-latitude and north polar ozone changes. □

Susan Solomon is at the NOAA Aeronomy Laboratory, Boulder, Colorado 80303-3328, USA.

1. National Research Council *Causes and Effects of Changes in Stratospheric Ozone: Update 1983* (National Academy Press, Washington, DC, 1984).
2. Molina, M. J. & Rowland, F. S. *Nature* **249**, 810–814 (1974).
3. Stolarski, R. A. & Cicerone, R. J. *Can. J. Chem.* **52**, 1610–1615 (1974).
4. Prinn, R. G. *et al.* *J. geophys. Res.* **88**, 8853–8868 (1983).
5. World Meteorological Organization, Rep. No. 20, Vol. 1 (1990).
6. Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
7. Wuebbles, D. J., Luther, F. M. & Penner, J. E. *J. geophys. Res.* **88**, 1444–1456 (1983).
8. Komhyr, W. D., Grass, R. D. & Leonard, R. K. *J. geophys. Res.* **93**, 1248–1251 (1988).
9. Komhyr, W. D., Oltmans, S. J. & Grass, R. D. *J. geophys. Res.* **93**, 5167–5184 (1988).
10. Chubachi, S. *Mem. Natl. Inst. Polar Res., Spec. Iss.* **34**, 13–18 (1984).
11. Andrews, D. G., Holton, J. R. & Leovy, C. B. *Middle Atmosphere Dynamics* (Academic, London, 1987).
12. Gardiner, B. G. & Shanklin, J. D. *J. geophys. Res.* **93**, 1199–1201 (1988).
13. Newman, P. A. *J. geophys. Res.* **93**, 1228–1231 (1988).
14. Callis, L. B. & Natarajan, M. *J. geophys. Res.* **91**, 10771–10780 (1986).
15. Tung, K. K., Ko, M. K. W., Rodriguez, J. M. & Sze, N. D. *Nature* **338**, 811–814 (1986).
16. Mahajan, J. D. & Fels, S. B. *J. geophys. Res.* **93**, 1316–1319 (1988).
17. Solomon, S., Garcia, R. R., Rowland, F. S. & Wuebbles, D. J. *Nature* **321**, 755–758 (1986).
18. McElroy, M. B., Salawitch, R. J., Wofsy, S. C. & Logan, J. A. *Nature* **321**, 759–762 (1986).
19. Solomon, S. *Rev. Geophys.* **26**, 131–148 (1988).
20. McKenzie, R. L. & Johnston, P. V. *J. geophys. Res.* **89**, 73–76 (1984).
21. Mount, G. H., Sanders, R. W., Schmeltekopf, A. L. & Solomon, S. *J. geophys. Res.* **92**, 8320–8328 (1987).
22. Fahey, D. W. *et al.* *J. geophys. Res.* **94**, 16665–16687 (1989).
23. Hofmann, D. J., Rosen, J. M. & Harder, J. W. *J. geophys. Res.* **93**, 605–616 (1988).
24. Podolske, J. R., Loewenstein, M., Strahan, S. E. & Chan, K. R. *J. geophys. Res.* **94**, 16767–16775 (1989).
25. McCormick, M. P., Steele, H. M., Hamill, P., Chu, W. P. & Swisler, T. J. *J. Atmos. Sci.* **39**, 1387–1397 (1982).
26. Molina, M. J., Tso, T. L., Molina, L. T. & Wang, F. C. Y. *Science* **238**, 1253–1257 (1987).
27. Tolbert, M. A., Rossi, M. J., Malhotra, R. & Golden, D. M. *Science* **238**, 1258–1260 (1987).
28. Leu, M. T. *J. geophys. Res.* **93**, 17–20 (1988).
29. Toon, O. B., Hamill, P., Turco, R. P. & Pinto, J. *J. geophys. Res.* **93**, 1284–1287 (1988).
30. Crutzen, P. J. & Arnold, F. *Nature* **324**, 651–655 (1986).
31. McElroy, M. B., Salawitch, R. J. & Wofsy, S. C. *J. geophys. Res.* **93**, 1296–1299 (1988).
32. Fahey, D. W. *et al.* *J. geophys. Res.* **94**, 11299–11316 (1989).
33. Kelly, K. K. *et al.* *J. geophys. Res.* **94**, 11317–11358 (1989).
34. Fahey, D. W. *et al.* *Nature* **344**, 321–324 (1990).
35. Toon, O. B., Turco, R. P. & Hamill, P. *J. geophys. Res.* **94**, 445–448 (1989).
36. Molina, L. T. & Molina, M. J. *J. Phys. Chem.* **91**, 433–436 (1987).
37. Brune, W. H., Anderson, J. G. & Chan, K. R. *J. geophys. Res.* **94**, 16639–16648 (1989).
38. Carroll, M. A., Sanders, R. W., Solomon, S. & Schmeltekopf, A. L. *J. geophys. Res.* **94**, 16633–16638 (1989).
39. de Zafra, R. L. *et al.* *Nature* **328**, 408–411 (1987).
40. Brune, W. H., Anderson, J. G. & Chan, K. R. *J. geophys. Res.* **94**, 16649–16665 (1989).

41. Sander, S. P., Friedl, R. R. & Yung, Y. L. *Science* **245**, 1095–1098 (1989).
42. Troler, M., Mauldin, R. L. III & Ravishankara, A. R. *J. phys. Chem.* **94**, 4896–5001 (1990).
43. Proffitt, M. H. *et al. J. geophys. Res.* **94**, 16547–16556 (1989).
44. Schoeberl, M. R. *et al. J. geophys. Res.* **94**, 16815–16846 (1989).
45. Yung, Y. L., Allen, M., Crisp, D., Zurek, R. W. & Sander, S. P. *Science* **248**, 721–724 (1990).
46. *J. geophys. Res.* **94**, Nos 9 and 14 (1989).
47. Barrett, J. W. *et al. Nature* **336**, 455–458 (1988).
48. Garcia, R. R. & Solomon, S. *J. geophys. Res.* **88**, 1378–1400 (1983).
49. Jones, R. L. *et al. J. geophys. Res.* **94**, 11529–11558 (1989).
50. Anderson, J. G., Brune, W. H., Toohey, D. W. & Proffitt, N. H. *Science* (in the press).
51. Rodriguez, J. M., Ko, M. K. W. & Sze, N. D. *Geophys. Res. Lett.* **13**, 1292–1295 (1986).
52. Solomon, S., Mount, G. H., Sanders, R. W. & Schmeltekopf, A. L. *J. geophys. Res.* **92**, 8329–8338 (1987).
53. Wahner, A. *et al. J. geophys. Res.* **94**, 11405–11411 (1989).
54. Farmer, C. B., Toon, G. C., Shaper, P. W., Blavier, J. F. & Lowes, L. L. *Nature* **329**, 126–130 (1987).
55. Coffey, M. T., Markin, W. G. & Goldman, A. *J. geophys. Res.* **94**, 16597–16614 (1989).
56. Toon, G. C. *et al. J. geophys. Res.* **94**, 16571–16596 (1989).
57. Anderson, J. G. *et al. J. geophys. Res.* **94**, 11480–11520 (1989).
58. McElroy, M. B. & Salawitch, R. J. *Planet. Space Sci.* **37**, 1653–1672 (1989).
59. Solomon, S., Sanders, R. W., Carroll, M. A. & Schmeltekopf, A. L. *J. geophys. Res.* **94**, 11393–11404 (1989).
60. Solomon, S., Sanders, R. W. & Miller, H. L. *J. geophys. Res.* (in the press).
61. Austin, J. *et al. J. geophys. Res.* **94**, 16717–16736 (1989).
62. Watterson, I. G. & Tuck, A. F. *J. geophys. Res.* **94**, 16511–16526 (1989).
63. Tuck, A. F. *J. geophys. Res.* **94**, 11687–11737 (1989).
64. Hartmann, D. L. *et al. J. geophys. Res.* **94**, 16779–16796 (1989).
65. McCormick, M. P., Zawodny, J. M., Veiga, R. E., Larsen, J. C. & Wang, P. H. *Planet. Space Sci.* **37**, 1567–1586 (1989).
66. Proffitt, M. H. *et al. J. geophys. Res.* **94**, 16797–16814 (1989).
67. Rodriguez, J. M., Ko, M. K. W. & Sze, N. D. *Geophys. Res. Lett.* **17**, 255–258 (1990).
68. Deshler, T., Hofmann, D. J., Hereford, J. V. & Sutter, C. B. *Geophys. Res. Lett.* **17**, 151–154 (1990).
69. Garcia, R. R. & Solomon, S. *Geophys. Res. Lett.* **14**, 848–851 (1987).
70. Ko, M. K. W. *et al. J. geophys. Res.* **94**, 16705–16716 (1989).
71. Blake, D. R. & Rowland, F. S. *Science* **239**, 1129–1131 (1988).
72. Angell, J. K. *J. Clim.* **1**, 1296–1304 (1988).
73. Trenberth, K. E. & Olson, J. G. *J. Clim.* **2**, 1196–1206 (1989).
74. Newman, P. A. & Randel, W. J. *J. geophys. Res.* **93**, 12585–12606 (1988).
75. Poole, L. R., Solomon, S., McCormick, M. P. & Pitts, M. C. *Geophys. Res. Lett.* **16**, 1157–1160 (1989).
76. Shine, K. P. *Geophys. Res. Lett.* **13**, 1331–1334 (1986).
77. Atkinson, R. J., Matthews, W. A., Newman, P. A. & Plumb, R. A. *Nature* **340**, 290–294 (1989).
78. Murphy, D. M. *et al. J. geophys. Res.* **94**, 11669–11686 (1989).
79. Tolbert, M. A., Rossi, M. J. & Golden, D. M. *Geophys. Res. Lett.* **15**, 851–853 (1988).
80. Mozurkewich, M. & Calvert, J. G. *J. geophys. Res.* **93**, 15889–15897 (1988).
81. Hofmann, D. J. & Solomon, S. *J. geophys. Res.* **94**, 5029–5042 (1989).
82. *Geophys. Res. Lett.* **17**, No. 4 (1990).
83. McKenna, D. S. *et al. Geophys. Res. Lett.* **17**, 553–556 (1990).
84. Nagatani, R. M., Miller, A. J., Gelman, M. E. & Newman, P. A. *Geophys. Res. Lett.* **17**, 333–336 (1990).

ACKNOWLEDGEMENTS. I thank D. W. Fahey, R. R. Garcia, G. Huebler, R. L. Jones, D. S. McKenna, D. M. Murphy, A. O'Neill, L. R. Poole, A. R. Ravishankara and A. F. Tuck for comments.

ARTICLES

Solid C₆₀: a new form of carbon

W. Krätschmer*, Lowell D. Lamb†, K. Fostiropoulos* & Donald R. Huffman†

*Max-Planck-Institut für Kernphysik, 6900 Heidelberg, PO Box 103980, Germany

†Department of Physics, University of Arizona, Tucson, Arizona 85721, USA

A new form of pure, solid carbon has been synthesized consisting of a somewhat disordered hexagonal close packing of soccer-ball-shaped C₆₀ molecules. Infrared spectra and X-ray diffraction studies of the molecular packing confirm that the molecules have the anticipated 'fullerene' structure. Mass spectroscopy shows that the C₇₀ molecule is present at levels of a few per cent. The solid-state and molecular properties of C₆₀ and its possible role in interstellar space can now be studied in detail.

FOLLOWING the observation that even-numbered clusters of carbon atoms in the range C₃₀–C₁₀₀ are present in carbon vapour¹, conditions were found^{2–4} for which the C₆₀ molecule could be made dominant in the large-mass fraction of vapourized graphite. To explain the stability of the molecule, a model was proposed of an elegant structure in which the carbon atoms are arranged at the 60 vertices of a truncated icosahedron, typified by a soccer ball. The structure, dubbed buckminsterfullerene² because of its geodesic nature, has been the subject of several theoretical stability tests^{5,6} and has been discussed widely in the literature. Calculations of many physical properties have been made, including electron energies^{7–9}, the optical spectrum⁹, vibrational modes^{10–15}, and the electric and magnetic properties^{16,17}. There has been speculation on the possible chemical and industrial uses of C₆₀ (ref. 2), and on its importance in astrophysical environments^{18–20}. Until now, it has not been possible to produce sufficient quantities of the material to permit measurement of the physical properties, to test the theoretical calculations, or to evaluate the possible applications.

Some of us have recently reported evidence^{21,22} for the presence of the C₆₀ molecule in soot condensed from evaporated

graphite. The identification was based primarily on the observed isotope shifts of the infrared absorptions when ¹²C was replaced by ¹³C, and on comparison of the observed features with theoretical predictions. The measured infrared and ultraviolet absorption bands were superimposed on a rather large continuum background absorption from the graphitic carbon which comprised ≥95% of the sample. Here we report how to extract the carrier of the features from the soot, how to purify it, and evidence that the material obtained is in fact primarily C₆₀.

Method of production

The starting material for our process is pure graphitic carbon soot (referred to below as simply soot) with a few per cent by weight of C₆₀ molecules, as described in refs 21, 22. It is produced by evaporating graphite electrodes in an atmosphere of ~100 torr of helium. The resulting black soot is gently scraped from the collecting surfaces inside the evaporation chamber and dispersed in benzene. The material giving rise to the spectral features attributed to C₆₀ dissolves to produce a wine-red to brown liquid, depending on the concentration. The liquid is then separated from the soot and dried using gentle heat, leaving a residue of dark brown to black crystalline material. Other non-polar solvents, such as carbon disulphide and carbon tetrachloride, can also dissolve the material. An alternative concentration procedure is to heat the soot to 400 °C in a vacuum or in an inert atmosphere, thus subliming the C₆₀ out of the soot (W. Schmidt, personal communication). The sublimed coatings are brown to grey, depending on the thickness. The refractive index in the near-infrared and visible is about two. To purify the material, we recommend removing the ubiquitous hydrocarbons before the concentration procedure is applied (for example, by washing the initial soot with ether). Thin films and powder samples of the new material can be handled without special precautions and seem to be stable in air for at least several weeks, although there does seem to be some deterioration with time for reasons that are as yet unclear. The material can be

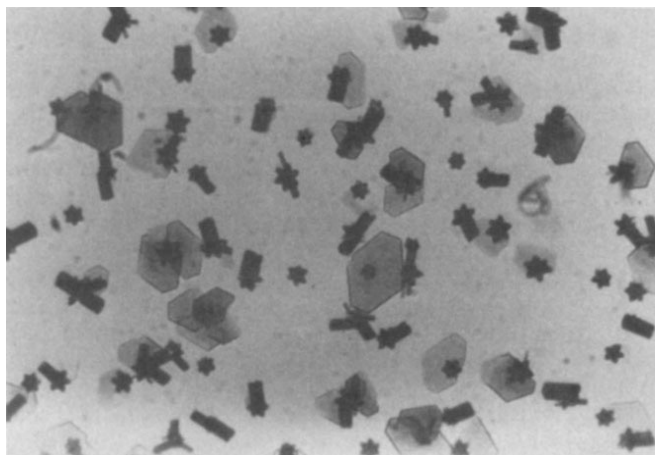


FIG. 1 Transmission micrograph of typical crystals of the C_{60} showing thin platelets, rods and stars of hexagonal symmetry.

sublimed repeatedly without decomposition. Using the apparatus described, one person can produce of the order of 100 mg of the purified material in a day.

Studies by optical microscopy of the material left after evaporating the benzene show a variety of what appear to be crystals—mainly rods, platelets and star-like flakes. Figure 1 shows a micrograph of such an assemblage. All crystals tend to exhibit six-fold symmetry. In transmitted light they appear red to brown in colour; in reflected light the larger crystals have a metallic appearance whereas the platelets show interference colours. The platelets can be rather thin and are thus ideally suited for electron-diffraction studies in an electron microscope (see the inset in Fig. 3).

Mass spectroscopy

The material has been analysed by mass spectrometry at several facilities. All mass spectra have a strong peak at mass 720 a.m.u., the mass of C_{60} . Significant differences in the spectra occur only at masses lower than 300 a.m.u. Most of these differences seem to originate from the different ionization techniques and in the different methods of desorbing molecules from the sample. Mass spectra recorded at low and high resolution are shown in Fig. 2. The spectra were obtained using a time-of-flight secondary-ion mass spectrometer²³ and a C_{60} -coated stainless-steel plate. In

the mass range above 300 a.m.u., the spectrum is dominated by C_{60} ions and its fragments (even-numbered clusters of atomic carbon), and C_{70} ions. In this sample, the ratio of C_{70} to C_{60} is ~ 0.1 . The high-resolution mass spectrum shows approximately the expected isotope pattern for C_{60} . The increasing background in the low-resolution mass spectrum is not produced by the sample—such backgrounds also occur in blank measurements on uncoated stainless-steel substrates.

So far, the cleanest mass spectra have been obtained when the material was evaporated and ionized in the vapour phase by electrons. In such spectra the low-mass background is substantially reduced and the entire mass spectrum is dominated by C_{60} ions and its fragments. The ratio of C_{70} to C_{60} in these mass spectra is ~ 0.02 and seems to be smaller than that shown in Fig. 2. Both ratios are of the order of those reported from laser-evaporation experiments^{2,3}. We assume, as previously suggested²⁴, that the C_{70} molecule also has a closed-cage structure, either elongated²⁴ or nearly spherical²⁵. Further details of the mass spectroscopy of the new material will be published elsewhere.

Structure

To determine if the C_{60} molecules form a regular lattice, we performed electron and X-ray diffraction studies on the individual crystals and on the powder. A typical X-ray diffraction pattern of the C_{60} powder is shown in Fig. 3. To aid in comparing the electron diffraction results with the X-ray results we have inset the electron diffraction pattern in Fig. 3. From the hexagonal array of diffraction spots indexed as shown in the figure, a d spacing of 8.7 Å was deduced corresponding to the (100) reciprocal lattice vector of a hexagonal lattice. The most obvious correspondence between the two types of diffraction is between the peak at 5.01 Å of the X-ray pattern and the (110) spot of the electron diffraction pattern, which gives a spacing of ~ 5.0 Å. Assuming that the C_{60} molecules are behaving approximately as spheres stacked in a hexagonal close-packed lattice with a c/a ratio of 1.633, d spacings can be calculated. The results are shown in Table 1. The values derived from this interpretation are $a = 10.02$ Å and $c = 16.36$ Å. The nearest-neighbour distance is thus 10.02 Å. For such a crystal structure the density is calculated to be 1.678 g cm^{-3} , which is consistent with the value of $1.65 \pm 0.05 \text{ g cm}^{-3}$ determined by suspending crystal samples in aqueous GaCl_3 solutions of known densities. Although the agreement shown in Table 1 is good, the absence of the characteristically strong (101) diffraction of the hexagonal close-packed structure, and the broad continuum in certain regions suggest that the order is less than

FIG. 2 Low-resolution (top) and high-resolution time-of-flight mass spectra of positive ions obtained from coatings of solid C_{60} . A 5-keV Ar^+ ion beam was used to sputter and ionize the sample. The isotope pattern (bottom) is approximately that expected for C_{60} molecules composed of ^{12}C and ^{13}C isotopes of natural abundance.

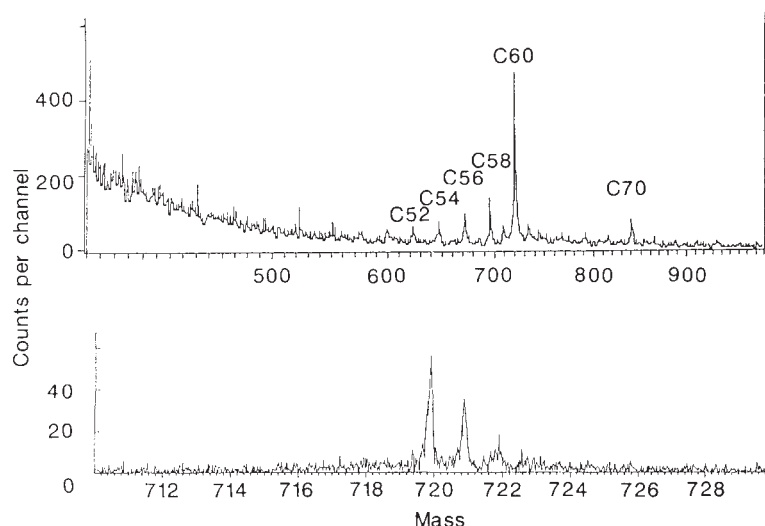


TABLE 1 X-ray diffraction results

Measured 2θ (deg)	Measured d spacing (Å)	Calculated d spacing (Å)	Assignment (hkl)
10.2 shoulder	8.7	8.68	(100)
10.81	8.18	8.18	(002)
		7.68	(101)
17.69	5.01	5.01	(110)
20.73	4.28	4.28	(112)
21.63	4.11	4.09	(004)
28.1	3.18	3.17	(114)
30.8	2.90	2.90	(300)
32.7	2.74	2.73	(006)

Assignments for a hexagonal lattice using $a=10.02$ Å, $c=16.36$ Å.
 $(1/d^2) = \frac{4}{3} [(h^2 + hk + k^2)/a^2] + l^2/c^2$.

perfect. Further, X-ray diffraction patterns from carefully grown crystals up to 500 μm in size with well developed faces yielded no clear spot pattern (in contrast to the electron diffraction pattern on micrometre-sized crystals). It therefore appears that these larger crystals do not exhibit long-range periodicity in all directions.

A likely explanation for these facts lies in the disordered stacking of the molecules in planes normal to the c axis. It is well known that the positions taken by spheres in the third layer of stacking determines which of the close-packed structures occurs, the stacking arrangement in a face-centred cubic structure being ABCABC... whereas that in a hexagonal close-packed structure is ABABAB... If the stacking sequence varies, the X-ray lines owing to certain planes will be broadened by the disorder whereas other lines will remain sharp. Such disordered crystalline behaviour was observed long ago in the hexagonal close-packed structure of cobalt²⁶⁻²⁸ where X-ray diffraction lines such as (101), (102) and (202) were found to be substantially broadened by the stacking disorder. Reflections from planes such as (002) remain sharp because these planes have identical spacings in the face-centred cubic and hexagonal close-packed structures. For the planes producing broadened diffraction peaks because of this kind of disorder, the following condition for the Miller indices (hkl) has been shown to apply^{27,29}: $h - k = 3t \pm 1$ (where t is an integer) and $l \neq 0$. None of these broadened reflections are apparent in the X-ray pattern of Fig. 3. This may explain the weakness of the characteristically strong (101) peak. Whether or not this stacking disorder is

related to the presence of the possibly elongated C_{70} molecule has yet to be determined.

In small crystals at least, the C_{60} molecules seem to assemble themselves into a somewhat ordered array as if they are effectively spherical, which is entirely consistent with the hypothesis that they are shaped like soccer balls. The excess between the nearest-neighbour distance (10.02 Å) and the diameter calculated for the carbon cage itself (7.1 Å) must represent the effective van der Waals diameter set by the repulsion of the π electron clouds extending outward from each carbon atom. Because the van der Waals diameter of carbon is usually considered to be 3.3–3.4 Å the packing seems a little tighter than one might expect for soccer-ball-shaped C_{60} molecules. The reason for this has not yet been determined.

In summary, our diffraction data imply that the substance isolated is at least partially crystalline. The inferred lattice constants, when interpreted in terms of close-packed icosahedral C_{60} , yield a density consistent with the measured value. Further evidence that the molecules are indeed buckminsterfullerene and that the solid primarily consists of these molecules comes from the spectroscopic results.

Spectroscopy

The absorption spectra of the graphitic soot^{21,22} showed evidence for the presence of C_{60} in macroscopic quantities. Following the purification steps described above the material can be studied spectroscopically with the assurance that the spectra are dominated by C_{60} , with some possible effects from C_{70} . Samples were prepared for spectroscopy by subliming pure material onto transparent substrates for transmission measurements. Depending on the pressure of helium in the sublimation chamber, the nature of the coatings can range from uniform films (at high vacuum) to coatings of C_{60} smoke (sub-micrometre microcrystalline particles of solid C_{60}) with the particle size depending to some extent on the pressure.

Figure 4 shows the transmission spectrum of an $\sim 2\text{-}\mu\text{m}$ -thick C_{60} coating on a silicon substrate. The infrared bands are at the same positions as previously reported^{21,22}, with the four most intense lines at 1,429, 1,183, 577 and 528 cm^{-1} ; here, however, there is no underlying continuum remaining from the soot. In many of our early attempts to obtain pure C_{60} , there was a strong band in the vicinity of 3.0 μm , which is characteristic of a CH-stretching mode. After much effort this contaminant was successfully removed by washing the soot with ether and using distilled benzene in the extraction. The spectrum in Fig. 4 was obtained when the material cleaned in such a manner was sublimed under vacuum onto the substrate. The spectrum shows

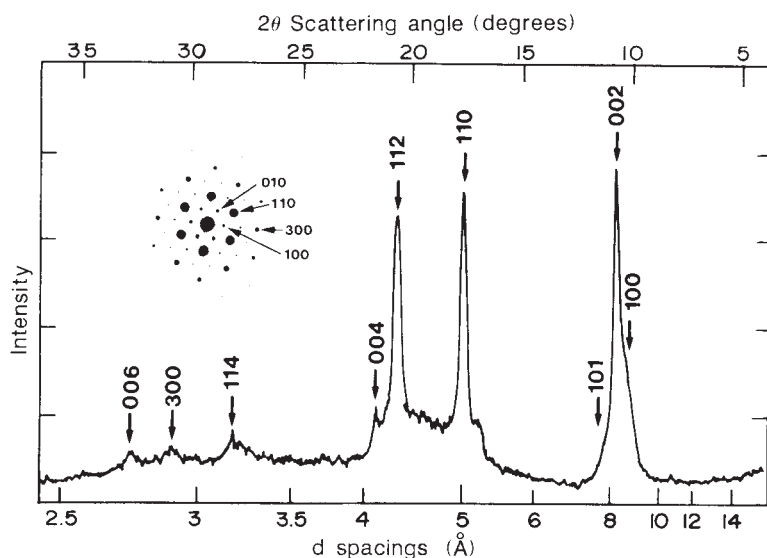


FIG. 3 X-ray diffraction pattern of a microcrystalline powder of C_{60} . Inset (upper left) is a single-crystal electron diffraction pattern indexed with Miller indices compatible with the X-ray pattern. The pattern is from a thin platelet such as those in Fig. 1 with the electron beam perpendicular to the flat face.

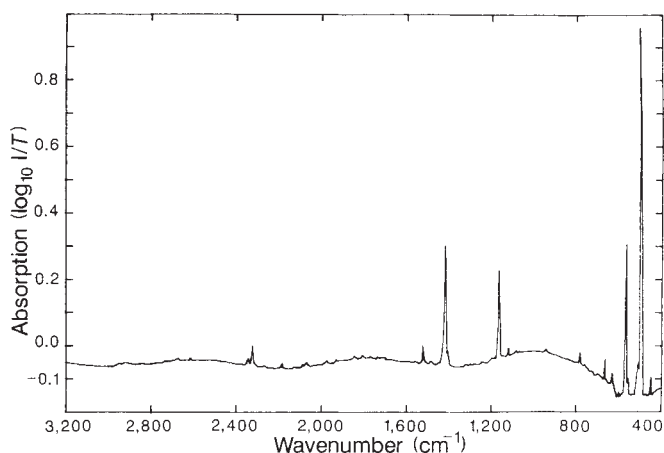


FIG. 4 Infrared absorption spectrum of a coating, $\sim 2 \mu\text{m}$ thick, of solid C_{60} on a silicon substrate, referenced to a clean silicon substrate. Apparent negative absorptions are due to the coating acting in part as a non-reflecting layer.

very little indication of CH impurities. Vibrational modes to compare with the measured positions of the four strong bands have been calculated by several workers^{10–15}. As noted previously, the presence of only four strong bands is expected for the free, truncated icosahedral molecule with its unusually high symmetry. Also present are a number of other weak infrared lines which may be due to other causes, among which may be absorption by the C_{70} molecule or symmetry-breaking produced (for example) by isotopes other than ^{12}C in the C_{60} molecule or by mutual interaction of the C_{60} molecules in the solid. Weaker features at $\sim 2,330$ and $2,190 \text{ cm}^{-1}$, located in the vicinity of the free CO_2 and CO stretching modes, may imply some attachment of the CO_2 or CO to a small fraction of the total number of C_{60} molecules. Another notable feature is the peak at 675 cm^{-1} , which is weak in the thin-film substrates but almost as strong as the four main features in the crystals. We suspect that this vibrational mode may be of solid state rather than molecular origin.

Figure 5 shows an absorption spectrum taken on a uniform film coated on a quartz glass substrate. The ultraviolet features are no longer obscured by the graphitic carbon background as in our previous spectra²². Broad peaks at 216, 264 and 339 nm dominate the spectra. Weaker structures show up in the visible, including a plateau between ~ 460 and 500 nm and a small peak near 625 nm . At the bottom of Fig. 5 we have shown positions

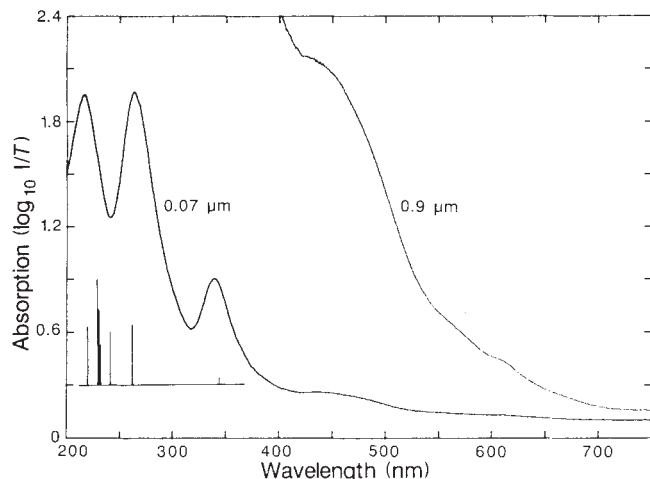


FIG. 5 Visible-ultraviolet absorption spectra of two thicknesses of solid C_{60} on quartz. The calculated⁹ positions and relative oscillator strengths for allowed transitions of C_{60} are shown on the bottom.

and relative oscillator strengths taken from Larsson, Volosov and Rosén⁹ calculated for the C_{60} molecule. They also reported a variety of forbidden bands with the lowest energy ones in the vicinity of 500 nm . There seems to be a rough correspondence between our measurements on solid films and the allowed transitions predicted for the molecule. The possibility exists, however, that one or more of the absorption features shown in Fig. 5 are due to C_{70} . We still do not observe a band at 386 nm in our films, as observed³⁰ using a laser depletion spectroscopy method and attributed to the C_{60} molecule. Quite similar spectra to that in Fig. 5 have been recorded for microcrystalline coatings deposited at helium pressures of 100 torr , for example. The peaks occur at the slightly shifted positions of 219 , 268 and 345 nm .

Possible interstellar dust

The original stimulus for the work² that led to the hypothesis of the soccer-ball-shaped C_{60} molecule, buckminsterfullerene, was an interest in certain unexplained features in the absorption and emission spectra of interstellar matter. These include an intense absorption band at 217 nm which has long been attributed to small particles of graphite³¹, a group of unidentified interstellar absorption bands in the visible that have defied explanation for more than 70 years^{31,32}, and several strong emission bands attributed to polycyclic aromatic hydrocarbons^{33,34}. Based on the visible and infrared absorption spectra of Figs 4 and 5, we do not see any obvious matches with the interstellar features. The ultraviolet band at $216\text{--}219 \text{ nm}$ has a similar peak wavelength to an interstellar feature, although the other strong bands of the spectrum have no interstellar counterparts. As the influence of C_{70} absorptions on the spectrum is not yet known, a conclusive comparison with the 217-nm interstellar band is difficult. We note that the visible-ultraviolet spectrum presented here is characteristic of a solid, rather than of free molecules. In addition, these new results do not relate directly to absorption in the free C_{60}^+ molecular ion, which has been envisaged¹⁹ to explain the diffuse interstellar bands. Nevertheless, these data should now provide guidance for possible infrared detection of the C_{60} molecule, if it is indeed as ubiquitous in the cosmos as some have supposed.

Summary

To our method for producing macroscopic quantities of C_{60} , we have added a method for concentrating it in pure solid form. Analyses including mass spectroscopy, infrared spectroscopy, electron diffraction and X-ray diffraction leave little doubt that we have produced a solid material that apparently has not been reported previously. We call the solid fullerite as a simple extension of the shortened term fullerene, which has been applied to the large cage-shaped molecules typified by buckminsterfullerene (C_{60}). The various physical and chemical properties of C_{60} can now be measured and speculations concerning its potential uses can be tested. □

Received 7 August; accepted 7 September 1990.

- Rohlfing, E. A., Cox, D. M. & Kaldor, A. *J. chem. Phys.* **81**, 3322–3330 (1984).
- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
- Zhang, Q. L. *et al. J. phys. Chem.* **90**, 525–528 (1986).
- Liu, Y. *et al. Chem. Phys. Lett.* **126**, 215–217 (1986).
- Newton, M. D. & Stanton, R. E. *J. Am. chem. Soc.* **108**, 2469–2470 (1986).
- Lüthi, H. P. & Almlöf, J. *Chem. Phys. Lett.* **135**, 357–360 (1987).
- Satpathy, S. *Chem. Phys. Lett.* **130**, 545–550 (1986).
- Haddon, R. C., Brus, L. E. & Raghavachari, K. *Chem. Phys. Lett.* **125**, 459–464 (1986).
- Larsson, S., Volosov, A. & Rosén, A. *Chem. Phys. Lett.* **137**, 501–504 (1987).
- Wu, Z. C., Jelski, D. A. & George, T. F. *Chem. Phys. Lett.* **137**, 291–294 (1987).
- Stanton, R. E. & Newton, M. D. *J. phys. Chem.* **92**, 2141–2145 (1988).
- Weeks, D. E. & Harter, W. G. *Chem. Phys. Lett.* **144**, 366–372 (1988).
- Weeks, D. E. & Harter, W. G. *J. chem. Phys.* **90**, 4744–4771 (1989).
- Elser, V. & Haddon, R. C. *Nature* **325**, 792–794 (1987).
- Slanina, Z. *et al. J. molec. Struct.* **202**, 169–176 (1989).
- Fowler, P. W., Lazzeretti, P. & Zanasi, R. *Chem. Phys. Lett.* **165**, 79–86 (1990).
- Haddon, R. C. & Elser, V. *Chem. Phys. Lett.* **169**, 362–364 (1990).
- Kroto, H. *Science* **242**, 1139–1145 (1988).
- Kroto, H. W. in *Polycyclic Aromatic Hydrocarbons and Astrophysics* (eds Léger, A. *et al.*) 197–206 (Reidel, Dordrecht, 1987).

20. Léger, A., d'Hendecourt, L., Verstraete, L. & Schmidt, W. *Astr. Astrophys.* **203**, 145–148 (1988).
21. Krätschmer, W., Fostropoulos, K. & Huffman, D. R. in *Dusty Objects in the Universe* (eds Bussoletti, E. & Clitstone, A. A.) (Kluwer, Dordrecht, in the press).
22. Krätschmer, W., Fostropoulos, K. & Huffman, D. R. *Chem. Phys. Lett.* **170**, 167–170 (1990).
23. Steffens, P., Niehuis, E., Friese, T. & Benninghoven, A. *Ion Formation from Organic Solids* (ed. Benninghoven, A.) Ser. chem. Phys. Vol. 25, 111–117 (Springer-Verlag, New York, 1983).
24. Kroto, H. W. *Nature* **329**, 529–531 (1987).
25. Schmalz, T. G., Seitz, W. A., Klein, D. J. & Hite, G. E. *J. Am. chem. Soc.* **110**, 1113–1127 (1988).
26. Hendricks, S. B., Jefferson, M. E. & Schultz, J. F. *Z. Kristallogr.* **73**, 376–380 (1930).
27. Edwards, O. S., Lipson, H. & Wilson, A. J. C. *Nature* **148**, 165 (1941).
28. Edwards, O. S. & Lipson, H. *Proc. R. Soc. A* **180**, 268–277 (1942).
29. Houska, C. R., Averbach, B. L. & Cohen, M. *Acta Metal.* **8**, 81–87 (1960).
30. Heath, J. R., Curl, R. F. & Smalley, R. E. *J. chem. Phys.* **87**, 4236–4238 (1987).
31. Huffman, D. R. *Adv. Phys.* **26**, 129–230 (1977).
32. Herbig, E. *Astrophys. J.* **196**, 129–160 (1975).
33. Léger, A. & Puget, J. L. *Astr. Astrophys. Lett.* **137**, L5–L8 (1984).
34. Allamandola, L. J., Tielens, A. G. & Barker, J. R. *Astrophys. J.* **290**, L25–L28 (1985).

ACKNOWLEDGEMENTS. W.K. and K.F. thank our colleagues F. Arnold, J. Kissel, O. Möhler, G. Natour, P. Sölter, H. Zscheeg, H. H. Eysel, B. Nuber, W. Kühlbrandt, M. Rentzea and J. Sawatzki. L.D.L. and D.R.H. thank our colleagues J. T. Emmert, D. L. Bentley, W. Bilodeau, K. H. Schramm and D. R. Luffer. D.R.H. thanks the Alexander von Humboldt Stiftung for a senior US Scientist award. We also thank H. W. Kroto and R. F. Curl for discussions.

Expression of cystic fibrosis transmembrane conductance regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells

Devra P. Rich, Matthew P. Anderson, Richard J. Gregory*, Seng H. Cheng*, Sucharita Paul*, Douglas M. Jefferson†, John D. McCann, Katherine W. Klinger‡, Alan E. Smith* & Michael J. Welsh§

Howard Hughes Medical Institute, Departments of Internal Medicine and Physiology and Biophysics, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA

* Genzyme Corporation and ‡ IG Laboratories Inc., One Mountain Road, Framingham, Massachusetts 01701, USA

† Department of Physiology, Tufts University School of Medicine, and Departments of Pediatrics and Medicine, New England Medical Center, Boston, Massachusetts 02111, USA

The cystic fibrosis transmembrane conductance regulator (CFTR) was expressed in cultured cystic fibrosis airway epithelial cells and Cl^- channel activation assessed in single cells using a fluorescence microscopic assay and the patch-clamp technique. Expression of CFTR, but not of a mutant form of CFTR (ΔF508), corrected the Cl^- channel defect. Correction of the phenotypic defect demonstrates a causal relationship between mutations in the CFTR gene and defective Cl^- transport which is the hallmark of the disease.

CYSTIC fibrosis (CF), the most common lethal genetic disease in Caucasians¹, is characterized by abnormal electrolyte transport in several organs, including lung, sweat gland, intestine and pancreas^{1,2}. Defective regulation of the apical membrane Cl^- channels that control the rate of transepithelial Cl^- transport is well documented in CF epithelia^{3–5}. In normal airway epithelia, Cl^- channels are activated (opened) by an increase in intracellular levels of cyclic AMP. Cyclic AMP opens Cl^- channels by activating cAMP-dependent protein kinase which is thought to phosphorylate either the Cl^- channel itself or an associated regulatory protein. As a result, the rate of transepithelial Cl^- secretion increases. In CF airway epithelia, Cl^- channels are present in the membrane but they cannot be activated by cAMP or cAMP-dependent protein kinase^{6,7}. As a result, CF epithelia fail to secrete Cl^- (and thereby fluid) towards the airway lumen. Defective Cl^- secretion probably contributes to the abnormal mucociliary clearance and lung disease that is the main cause of morbidity and mortality in patients with CF.

The CFTR gene sequence was recently identified and shown to be mutated in patients with CF^{8–10}. A 3-base-pair deletion

resulting in the loss of a phenylalanine residue at position 508 (ΔF508) occurs in ~70% of CF chromosomes^{10,11}. In the accompanying manuscript¹² we describe construction of a full-length CFTR coding sequence. By *in vitro* and *in vivo* synthesis we have shown that CFTR associates with membranes, is a glycoprotein and can be phosphorylated; these properties are consistent with those predicted from the DNA sequence⁹. This work enabled us to test the hypothesis that expression of normal CFTR will complement the CF phenotype and should eventually allow an assessment of the function of CFTR.

Expression of CFTR in CF cells

In evaluating expression systems, we had two main considerations: the protein product should be processed appropriately and most of the cells should express the protein. We chose the vaccinia-T7 hybrid expression system developed by Moss and co-workers^{13,14} because it has several advantages, including: (1) infection and expression occur in a wide variety of eukaryotic cells; (2) the virus provides the enzymes necessary for transcription in the cytoplasm of the infected cell, thereby eliminating the need for nuclear processing; (3) the gene product can be correctly processed and targeted to the cell surface; and (4) a high percentage of cells express the recombinant DNA. We were also encouraged to adopt this system because it has been successfully used to express a functional K^+ channel in several cell types¹⁵, including epithelial cells (J.D.McC. and M.J.W., unpublished observations).

In the accompanying manuscript¹² we showed that HeLa cells produced CFTR when they were infected with a recombinant vaccinia virus expressing bacteriophage T7 RNA polymerase, and subsequently transfected with a plasmid containing the T7 promoter and terminator sequences flanking the CFTR coding sequence. We transfected a CF airway epithelial cell line (JME/CF15)¹⁶ with constructs containing either the normal CFTR coding sequence (pTM-CFTR-3), or the mutated CFTR coding sequence (pTM-CFTR-3 ΔF508). Figure 1a shows expression of CFTR in JME/CF15 cells at 6.5 hours after

§ To whom correspondence should be addressed.

transfection. At that time, vaccinia virus has turned off most host cell protein synthesis and directs synthesis primarily of proteins needed for viral replication¹⁷. CFTR migrates as a doublet of relative molecular mass 130,000–135,000 (M_r 130–135K); this doublet probably reflects differences in the state of glycosylation¹². The maximal rate of CFTR synthesis occurred 6.5–9.5 h after transfection. Figure 1b shows that both normal and $\Delta F508$ mutant CFTR are produced by the vaccinia virus expression system in transformed cultures of CF airway epithelial cells. CFTR and CFTR $\Delta F508$ have the same M_r .

Successful expression of CFTR in CF epithelial cells provided

us with a system in which to test the effect of the protein on regulation of Cl^- channels.

Assessment of Cl^- permeability

We used the vaccinia virus system in an attempt to express the gene in a high proportion of cells. But, because expression is dependent both on infection with the recombinant vaccinia virus and on subsequent transfection with the CFTR-containing plasmid, it was possible that only a fraction of the cells would express the functional protein. In such a case it would be necessary to assay Cl^- transport in

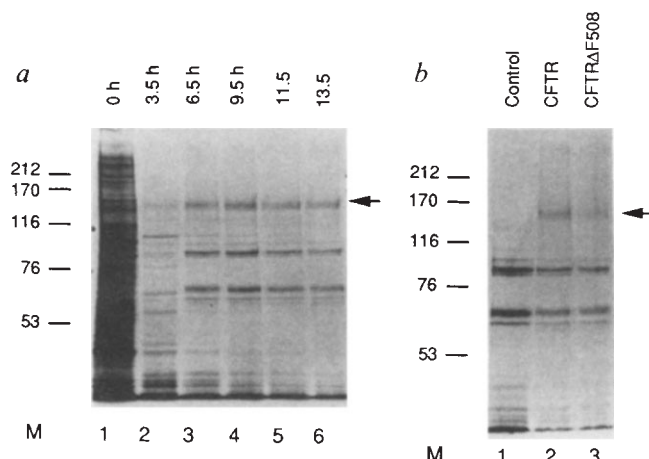


FIG. 1 Expression of CFTR in CF airway epithelial cells. *a*, Transformed CF airway epithelial cells (JME/CF15) were infected with recombinant vaccinia virus vTF-3 (to provide cytoplasmic expression of the bacteriophage T7 RNA polymerase, ref. 13) and transfected with pTM-CFTR-3 (ref. 12). Cells were collected at the indicated times after transfection. Cells in lane 1 (labelled 0 h) were not infected or transfected. Cells were metabolically labelled with [³⁵S]methionine for 1 h before collection. In *a* and *b*, cell-lysate proteins were separated by electrophoresis on a 7.5% SDS-polyacrylamide gel and [³⁵S]methionine-labelled proteins were detected by autoradiography. A doublet at 130–135K is clearly apparent at 6.5 h and beyond. *b*, Expression of CFTR and CFTR $\Delta F508$ in cultures of CF airway epithelial cells (JME/CF15). Cells were infected with recombinant vaccinia virus vTF-3 and transfected with either a control plasmid pTM1 (lane 1), pTM-CFTR-3 (lane 2), or pTM-CFTR-3 $\Delta F508$ (lane 3)¹². Twelve hours after transfection, cells were metabolically labelled with [³⁵S]methionine for 1 h and collected. Identical results were obtained on transformation of primary cultures of CF nasal epithelia. Molecular weight markers (K) are indicated in lane M. The arrow designates normal and mutant CFTR protein.

METHODS. Cell culture: cells were either primary cultures of normal airway epithelial cells (three preparations), primary cultures of CF airway epithelial cells (five preparations), or CF airway epithelial cells transformed by the SV40 large T antigen (JME/CF12 and JME/CF15), which retain the CF phenotype¹⁶. JME/CF15 is homozygous for the $\Delta F508$ mutation and JME/CF12 is heterozygous for the $\Delta F508$ mutation. Cells were cultured as previously described^{16,33}. Infection-transfection: Cells were plated at $\sim 5 \times 10^4$ cells cm^{-2} in 35-mm dishes 24 h before infection. Recombinant vaccinia virus vTF-3 was added to the cells for 1 h in serum-free medium at a high multiplicity of infection (10–20 m.o.i.). Cells were then transfected with recombinant plasmids (5 μg plasmid per 10^6 cells) using lipofectin (BRL)³⁴ (20 μg lipid per 10^6 cells) and incubated at 37 °C. Metabolic labelling: Cells were labelled with [³⁵S]methionine (25 $\mu\text{Ci ml}^{-1}$) for 1 h in hypertonic medium (190 mM NaCl) following a 30-min incubation in the absence of methionine. Cells were washed three times with ice-cold phosphate-buffered saline to remove excess label and collected in 250 μl of 25 mM 4-morpholinepropane sulphonic acid (pH 8.0), 2.5% 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propane sulphate, 185 mM KCl, 1 mM (ethylenedinitrilo)-tetraacetic acid, 100 $\mu\text{g ml}^{-1}$ aprotinin, and 17 μM PMSF. Cell lysate (20 μl) was added to an equal volume of SDS-sample buffer and analysed by SDS-polyacrylamide gel electrophoresis³⁵ and autoradiography. Gels were exposed to film for 2.5 days (*a*) and 5 days (*b*).

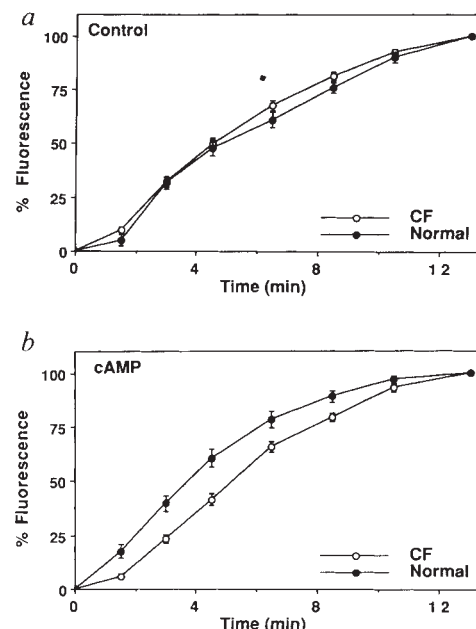


FIG. 2 Change in fluorescence of SPQ-loaded cells following substitution of NO_3^- for Cl^- in the bathing media. Comparison of normal and CF cells under control (*a*) or stimulated (20 μM forskolin) (*b*) conditions. Data are mean $\pm$ s.e.m. for 70 control and 58 stimulated normal cells (two different primary cultures) and 61 control and 51 stimulated CF cells (two different primary cultures). Values for stimulated normal cells were significantly greater than values for CF cells under either condition and values for normal cells under control conditions ($P < 0.05$).

METHODS. SPQ fluorescence was assayed 7–14 h after transfection. Cells were grown on glass coverslips and loaded with SPQ by including 10 mM SPQ in the medium for 12–18 h before use. SPQ fluorescence was initially quenched by incubating cells for 25–45 min in a buffer containing 135 mM NaI, 2.4 mM K_2HPO_4 , 0.6 mM KH_2PO_4 , 10 mM HEPES buffer (pH 7.4), and 10 mM dextrose. For some studies the buffer contained 1 mM MgSO_4 and 1 mM CaSO_4 , for others it contained 3 mM MgSO_4 and 1 mM ethylene glycol-bis (aminoethyl ether) N,N,N',N' -tetraacetic acid; similar results were obtained with both. Isoproterenol (2 μM) or forskolin (20 μM) were added as indicated, 5 min before the anion substitution. Indomethacin (2–3 μM) was included in all solutions for studies that compared control with stimulated cells. After measuring fluorescence for at least 2 min, the 135 mM NaI solution was replaced by one containing 135 mM NaNO_3 (time zero) and fluorescence was measured for another 13 min. Fluorescence of SPQ in single cells was measured using a Nikon inverted microscope, a SPEX digital imaging system, and a DAGE SIT66 camera. Excitation was at 350 nm and emission was at > 410 nm. Studies were at room temperature. Examination of Fig. 3a and b gives an indication of the experimental data. Cells were chosen for quantitation of fluorescence from a bright field image or from a single fluorescence image without knowledge of the rate of change. To correct for cell-to-cell differences in absolute fluorescence intensity (due to differences in cellular concentration of SPQ and differences in cell size), we present the data as '% Fluorescence' which is $100 \times (F_t - F_0) / (F_{13} - F_0)$, where F_0 , F_{13} and F_t are fluorescence intensity at 0 min, 13 min and at time t , respectively. Statistical significance between groups of averaged data in Figs 2–4 was assessed by t -test at the 3-min time point.

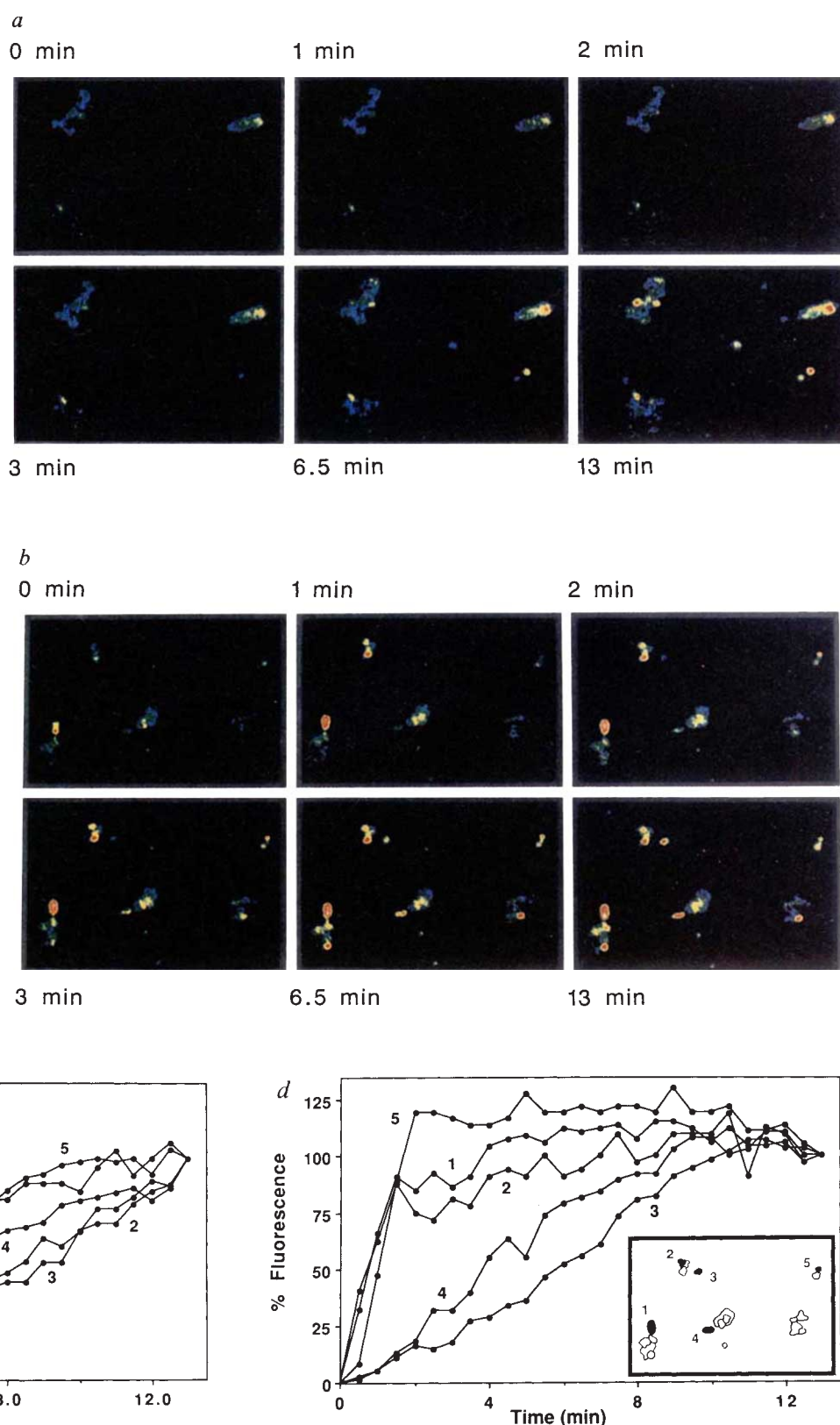
individual cells. Consequently we decided to use a fluorescence microscopic assay because it is well suited to the assessment of Cl^- permeability in a large number of individual cells.

The rapidly responding halide-sensitive fluorophore 6-methoxy-*N*-(3-sulphopropyl) quinolinium (SPQ) has been used to measure relative intracellular Cl^- concentration in both sub-

cellular vesicles and cells¹⁸⁻²⁰. It has been used to assess a cAMP-activated Cl^- channel in cultured canine tracheal epithelial cells²⁰, and Cl^- permeability in cultured cells from the sweat gland duct where it was able to detect the difference in Cl^- permeability between normal and CF cells¹⁹.

We have exploited the fact that the fluorescence of SPQ is more effectively quenched by I^- than by Cl^- or Br^- (ref. 18).

FIG. 3 Expression of CFTR increases halide permeability of CF airway epithelial cells. *a* and *b*, Pseudo-colour images of primary cultures of CF cells transfected with either pTM-CFTR-3ΔF508 (*a*) or pTM-CFTR-3 (*b*). Colour corresponds to fluorescence intensity: blue is the lowest, then yellow, red, and white is the highest. In this and all subsequent studies, cells from the same culture were simultaneously transfected with pTM-CFTR-3 or pTM-CFTR-3ΔF508. All solutions contained isoproterenol (2 μM) added 5 min before the ion substitution. The images shown were obtained just before substituting NaNO_3 for NaI (time 0) and 1, 2, 3, 6.5 and 13 min after the substitution, as indicated. The width of each field is 340 μm. *c* and *d*, Changes in fluorescence intensity for five of the cells shown in *a* and *b*, respectively; each of the series of data points and lines was obtained from a different cell. The insets identify the cells for which intensity is plotted; these cells were deliberately chosen to represent the heterogeneity observed in the cell population. *e*, Average values for cells transfected with pTM-CFTR-3ΔF508 ($n=72$ cells) or pTM-CFTR-3 ($n=65$ cells). Data are mean $\pm$ s.e.m. from two different primary CF cell cultures and two transformed cell lines; there was no difference between primary cultures and the cell lines. Values for CFTR cells were significantly greater than those for CFTRΔF508 cells ($P<0.05$). *f*, Histogram showing the distribution of cells that had undergone a given change in fluorescence by 4.5 min after substitution of NaNO_3 for NaI . Crosshatched bars indicate cells transfected with pTM-CFTR-3 ($n=65$); solid bars indicate cells transfected with pTM-CFTR-3ΔF508 ($n=72$). Cultures of CF cells were simultaneously infected/transfected with pTM-CFTR-3 or pTM-CFTR-3ΔF508 and studied 7–13 h later. All experiments and data analysis for transfected CF cells were performed blind: cells were coded and the experimenter had no knowledge of their identity except when this became apparent through experimental observation.



The cAMP-activated Cl^- channel is permeable to I^- , whereas many other Cl^- transport processes do not carry I^- (refs 21, 22). As a result, measurement of I^- flux across the cell membrane provides a more selective assay of the Cl^- channel than measurement of Cl^- or Br^- flux²¹. In preliminary studies with canine tracheal epithelial cells, we quenched intracellular SPQ fluorescence by incubating cells in buffered NaI. When the NaI was replaced by NaNO_3 , fluorescence intensity increased; the Cl^- channel conducts NO_3^- (ref. 22), but NO_3^- neither quenches SPQ nor competes with the I^- quenching of SPQ (ref. 23; and M.J.W., unpublished observations).

Figure 2a shows that under control conditions, the rate of change in fluorescence for primary cultures of normal human airway epithelial cells was the same as that for primary cultures of CF cells. But, when forskolin was added to increase cellular levels of cAMP, normal cells had a greater rate of fluorescence increase than CF cells (Fig. 2b), suggesting that forskolin activated Cl^- channels in normal but not in CF cells. Similar effects were obtained with isoproterenol. These results are consistent with previous studies demonstrating defective regulation of Cl^- channels in CF cells (reviewed in refs 2 and 5). These data indicate that the SPQ fluorescence assay can be used to test for complementation of the Cl^- transport defect in CF cells.

Cl^- permeability in CF cells

To determine whether CFTR corrects the Cl^- permeability defect in CF cells, we expressed CFTR in primary cultures and transformed CF airway epithelial cells, expressing CFTR Δ F508 as a control. Cells were bathed in NaI solution which was replaced by NaNO_3 at time zero; both solutions contained isoproterenol ($2 \mu\text{M}$) to activate Cl^- channels. Figure 3a–d shows the results

of two representative experiments and Fig. 3e shows the average data for all experiments.

There are two main observations (Fig. 3). First, and most importantly, the rate of fluorescence increase was greater in CF cells transfected with pTM-CFTR-3 than in those transfected with pTM-CFTR-3 Δ F508. Similar results were obtained when comparisons were made between cells transfected with pTM-CFTR-3 and the parent plasmid containing no complementary DNA (data not shown). These observations indicate that expression of CFTR, but not CFTR Δ F508, increase Cl^- permeability in CF cells. Second, there was cell-to-cell variability in the response which is apparent from Fig. 3a and b. Some of the cells transfected with pTM-CFTR-3 had a high permeability to halide (undergoing a maximal change in fluorescence 4.5 min after the ion substitution) (Fig. 3f), whereas halide permeability in other cells was comparable to that in cells transfected with pTM-CFTR-3 Δ F508. Such heterogeneity may result from cell-to-cell variability in infection or transfection or in the level of expression of the functional protein. The heterogeneous response means that averaged data (such as those shown in Fig. 3e and subsequently in Fig. 4) underestimate the effect of CFTR expression on Cl^- channels.

Comparison of Figs 3e and 2b suggests that under stimulated conditions, the rate of change in fluorescence is greater in CF cells expressing CFTR than in normal cells. The difference is even more striking when the heterogeneous nature of the response in CFTR-transfected cells is considered (note that Fig. 3e shows the average for all the cells). The two groups of cells probably express very different amounts of CFTR; endogenous levels of CFTR are expected to be low in primary cultures of airway cells, on the basis of very low levels of messenger RNA^{9,16}, whereas the transfected cells produce a large amount of CFTR (Fig. 1). Therefore we speculate that the amount of CFTR may determine the Cl^- permeability of the cell. This speculation is consistent with previous studies^{24,25} measuring the rate of sweat secretion in response to agents that increase cellular levels of cAMP. Obligate heterozygotes for mutations in the CFTR gene (parents of patients with CF) had a rate of sweat secretion that was intermediate between that observed in normal individuals and in patients with CF, implying a dosage relationship between CFTR and Cl^- secretion.

Figure 4 shows that isoproterenol increased the rate of fluorescence change in cells transfected with pTM-CFTR-3 but not with pTM-CFTR-3 Δ F508. This observation indicates that expression of CFTR restores cAMP-dependent changes in halide permeability in CF cells. We also noted that even under control conditions, CFTR-expressing cells had a significantly greater permeability to halide than cells expressing CFTR Δ F508 (compare Fig. 4a and b). We do not know the explanation for this unexpected difference, but it may mean that some of the CFTR is active at basal levels of cAMP.

cAMP-activated Cl^- current in CF cells

The fact that many (perhaps 40–60%) of the CFTR-transfected CF cells had an increased permeability to halide encouraged us to attempt the more difficult patch-clamp technique on single transfected cells. The advantage of the patch-clamp technique is that it provides a more direct assay of Cl^- channel currents. We used the perforated-patch, whole-cell recording technique^{26,27} in which the sum of current flow through all the channels in a single cell is recorded. Nystatin, when included in the patch pipette, electrically couples the pipette solution to the cell interior, and allows the exchange of monovalent ions between the pipette and the cytosol while preventing dialysis of other intracellular components. Thus it prevents the decay of cAMP-activated currents observed in many cells studied with the conventional whole-cell patch-clamp technique^{27,28}. Furthermore, when the Cl^- concentration in the pipette solution approximates the intracellular Cl^- concentration, it prevents large spontaneous changes in currents observed with the conven-

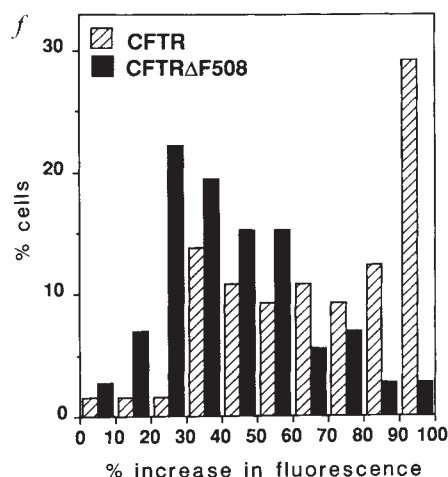
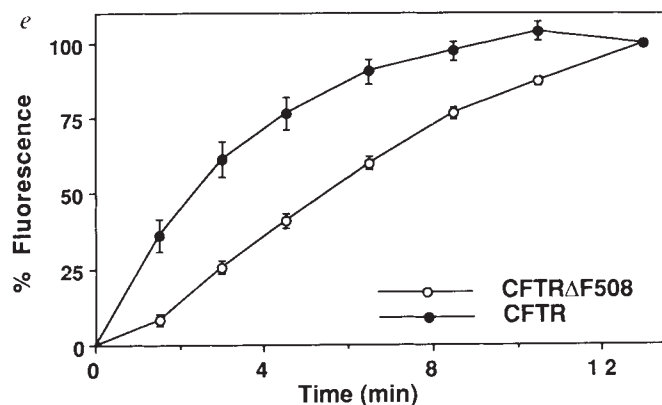


TABLE 1 Comparison of currents

	N	Baseline (pA)	cAMP (ΔpA)	Ca ²⁺ (ΔpA)
CFTRΔF508	6	54 ± 17	+13 ± 4	+945 ± 373
CFTR				
Responders	4	68 ± 20	+173 ± 95*	ND
Nonresponders	2	40	-3	ND

Comparison of currents (mean ± s.e.m.) measured at +80 mV under baseline conditions and in the presence of 1 mM CPT-cAMP (cAMP) or 5 mM CaCl₂ and 0.2 μM ionomycin (Ca²⁺). ΔpA, change in current in pA.

* Value significantly different from value obtained in cells transfected with pTM-CFTR-3ΔF508 ($P < 0.02$ by one-tailed *t*-test). ND, Not determined.

tional whole-cell technique (refs 29, 30; and M.P.A. and M.J.W., unpublished observations).

Figure 5 shows representative current tracings from CF cells transfected with pTM-CFTR-3ΔF508 (Fig. 5a and c) and pTM-CFTR-3 (Fig. 5b and d). Cyclic AMP fails to change the current in cells transfected with pTM-CFTR-3ΔF508, a result consistent with previous observations that cAMP does not activate CF Cl⁻ channels^{6,7}. Although none of the pTM-CFTR-3ΔF508-transfected cells respond to cAMP, they all subsequently show an increase in Cl⁻ current when the intracellular Ca²⁺ concentration is increased by adding Ca²⁺ and ionomycin (Fig. 5a, c, and Table 1). The Ca²⁺-activated Cl⁻ current showed time-dependent activation at depolarizing voltages, a characteristic of Ca²⁺-activated Cl⁻ currents in human airway epithelial cells (M.P.A. and M.J.W., unpublished observations) and in a Cl⁻-secreting intestinal epithelial cell line (T84) (ref. 30; and M.P.A. and M.J.W., unpublished observations). Subsequent addition of the Cl⁻ channel blocker, diphenylamine-2-carboxylate (DPC)^{29,31}, inhibited the Cl⁻ current.

In contrast to cells transfected with pTM-CFTR-3ΔF508, cAMP stimulated a Cl⁻ current in cells transfected with pTM-CFTR-3 (Fig. 5b, d, and Table 1). Cyclic AMP-dependent stimulation occurred after a short delay and was maximal within 1.5–4.5 min. We have observed a similar delay in the stimulation of cAMP-activated Cl⁻ channel current in nontransfected normal airway epithelial cells³ and in T84 cells (M.P.A. and M.J.W., unpublished observations); the delay could reflect the time necessary for agonist diffusion and for subsequent phosphoryla-

tion of target proteins. The SPQ studies suggested that not all the cells would respond to cAMP, and in fact we found that two of the six cells failed to increase current after addition of cAMP (Table 1). Addition of DPC inhibited the current (Fig. 5d).

Evidence that the cAMP-activated current is a Cl⁻ current includes: (1) the lack of significant voltage-dependence is characteristic of cAMP-activated whole-cell Cl⁻ currents in T84 and airway epithelial cells (ref. 30; and M.P.A. and M.J.W., unpublished observations); (2) the currents were outwardly rectifying and Cl⁻ is the only ion present that is likely to carry outward current (outward K⁺ currents are inhibited because Na⁺ exchanges with intracellular K⁺ through the nystatin pores^{26,27} and outward Na⁺ currents are inhibited by amiloride); (3) the Cl⁻ channel blocker DPC inhibited the current (DPC does not acutely inhibit Na⁺ or K⁺ channels in these cells; ref. 31, and M.J.W., unpublished observations); (4) the reversal potentials for cAMP-activated and Ca²⁺-activated Cl⁻ currents were the same and with ion substitutions they shifted in the direction expected for current flowing through anion-selective channels.

Discussion

CFTR was originally identified as the gene responsible for CF on the basis of several independent observations: the CFTR gene mapped to the correct chromosomal region; there was an appropriate pattern of tissue expression; and a mutation in CFTR was identified that is present on 70% of CF chromosomes and absent on non-CF chromosomes^{9–11}. Our data show that expression of CFTR corrects the CF defect in Cl⁻ transport, thereby demonstrating a causal relationship between mutations in the CFTR gene and the CF phenotype. Because expression of CFTRΔF508 did not complement the CF phenotype, other interpretations are very unlikely.

The data also strengthen the relationship between CF and epithelial Cl⁻ channels. Defective phosphorylation-dependent regulation of Cl⁻ channels has been the best characterized biochemical abnormality in CF. The fact that expression of CFTR corrects this abnormality suggests that CFTR is either a Cl⁻ channel or that it regulates Cl⁻ channels. It is unlikely that CFTR corrects a defect in channel regulation by affecting the synthesis of Cl⁻ channels or by affecting the processing of Cl⁻ channels during synthesis, because the vaccinia virus system inhibits synthesis of cellular proteins. Although our results can-

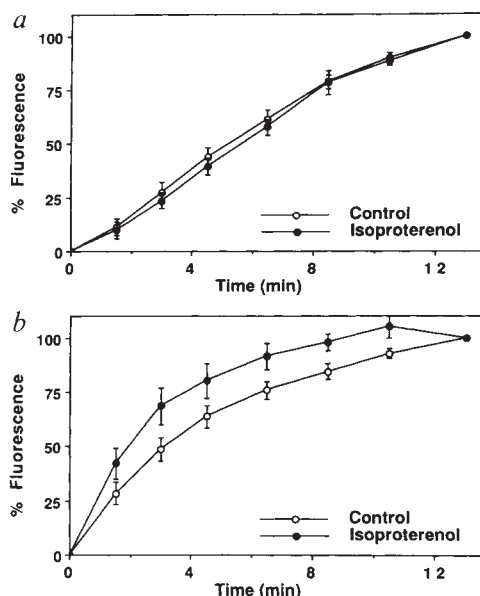


FIG. 4 Effect of stimulation on the response of CF cells transfected with either pTM-CFTR-3ΔF508 (a) or pTM-CFTR-3 (b). Experiments were performed as described in legends of Figs 2 and 3. Data are from 34 cells transfected with pTM-CFTR-3ΔF508 and 32 cells transfected with pTM-CFTR-3 from three preparations originating from two different primary cultures and one preparation of JME/CF 15 cells. Values for isoproterenol-treated CFTR cells were significantly greater than for control CFTR cells. Values from both groups of CFTR cells were different from those obtained in cells expressing CFTRΔF508 ($P < 0.05$).

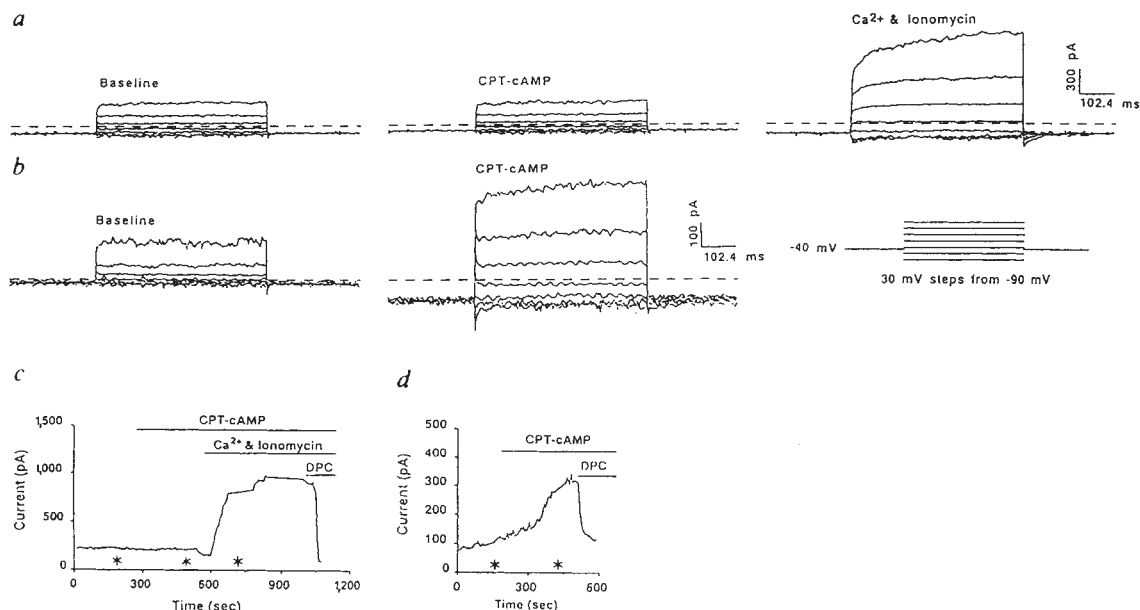


FIG. 5 Expression of CFTR induces cAMP-activated Cl^- currents in CF airway epithelial cells. Examples are shown from primary cultures of CF cells transfected with either pTM-CFTR-3 Δ F508 (*a* and *c*) or pTM-CFTR-3 (*b* and *d*). *a* and *b*, Tracings of currents obtained under baseline conditions and in the presence of 1 mM 8-(4-chlorophenylthio)-cAMP (CPT-cAMP) and 0.2 μM ionomycin and 2 mM CaCl_2 . The insets show the voltage protocol for the current tracings and the time and current scales. *c* and *d*, Time course of current measured from the cells used for *a* and *b*; values were obtained at +80 mV. The current tracings were obtained at the points marked by the asterisks. CPT-cAMP, ionomycin and Ca^{2+} , and DPC (3 mM) were present during the indicated times.

METHODS. Currents were measured using the perforated-patch modification²⁶ of the whole-cell patch-clamp technique^{27–29}. All cells were primary cultures of CF airway epithelia (four separate cultures from two different

subjects). Pipette tips were filled with a solution containing: 110 mM sodium gluconate, 25 mM NaCl, 8 mM MgCl_2 and 10 mM HEPES buffer (pH 7.3 with NaOH). Pipettes were then backfilled with the same solution containing nystatin (0.36 mg ml^{-1} ; Sigma). The external bath solution contained 135 mM NaCl, 2.4 mM K_2HPO_4 , 0.6 mM KH_2PO_4 , 3 mM MgCl_2 , 1 mM sodium EGTA, 10 mM HEPES (pH 7.4 with NaOH), and 10 mM dextrose. Amiloride (10 μM) and indomethacin (3 μM) were added to the bath solution to block epithelial Na^+ channels and cyclooxygenase, respectively. Cells were incubated in bath solution for 0.5–2 h at 25 °C before each experiment, to minimize baseline currents. Membrane voltage was maintained at –40 mV and depolarized to +80 mV for 500 ms every 3 s. Series resistance and cell capacitance were compensated as previously described²⁹. Cells were studied at 30–35 °C, 9–14 h after transfection.

not distinguish whether CFTR is itself a Cl^- channel or functions to regulate such channels, it will enable us to investigate the function of the protein. The relation between CFTR function and other cellular abnormalities such as increased Na^+ absorption³² must await future studies.

Our observation that overexpression of CFTR Δ F508 does not

correct the Cl^- transport abnormality suggests that a therapeutic strategy based on enhancing the expression of the mutant gene is unlikely to succeed. But demonstration that the CF phenotype can be corrected in cultured CF cells suggests the feasibility of a therapeutic approach based on correction of the underlying defect. □

Received 27 July; accepted 6 September 1990.

- Boat, T. F., Welsh, M. J. & Beaudet, A. L. in *The Metabolic Basis of Inherited Disease* (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 2649–2860 (McGraw Hill, New York, 1989).
- Quinton, P. *FASEB J.* **4**, 2709–2717 (1990).
- Welsh, M. J. & Liedtke, C. M. *Nature* **322**, 467–470 (1986).
- Frizzell, R. A., Reckemmer, G. & Shoemaker, R. L. *Science* **233**, 558–560 (1986).
- Welsh, M. J. *FASEB J.* **4**, 2718–2725 (1990).
- Schoumacher, R. A. et al. *Nature* **330**, 752–754 (1987).
- Li, M. et al. *Nature* **331**, 358–360 (1988).
- Rommens, J. M. et al. *Science* **245**, 1059–1065 (1989).
- Riordan, J. R. et al. *Science* **245**, 1066–1073 (1989).
- Kerem, B.-S. et al. *Science* **245**, 1073–1080 (1989).
- Lemna, W. K. et al. *New Engl. J. Med.* **322**, 291–296 (1990).
- Gregory, R. J. et al. *Nature* **347**, 382–386 (1990).
- Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8122–8126 (1986).
- Elroy-Stein, O., Fuerst, T. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6126–6130 (1989).
- Leonard, R. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7629–7633 (1989).
- Jefferson, D. M. et al. *Am. J. Physiol.* (in the press).
- Moss, B. in *Virology* (eds Fields, B. N. et al.) 685–703 (Raven, New York, 1985).
- Illsley, N. P. & Verkman, A. S. *Biochemistry* **26**, 1215–1219 (1987).
- Ram, S. J. & Kirk, K. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 10166–10170 (1989).
- Chao, A. C., Widdicombe, J. H. & Verkman, A. S. *J. Membrane Biol.* **113**, 193–202 (1990).
- Clancy, J. P., McCann, J. D., Li, M. & Welsh, M. J. *Am. J. Physiol.* **258**, L25–32 (1990).

- Li, M., McCann, J. D. & Welsh, M. J. *Am. J. Physiol.* **259**, C295–C301 (1990).
- Verkman, A. S., Sellers, M. C., Chao, A. C., Leung, T. & Ketcham, R. *Analyt. Biochem.* **178**, 355–361 (1989).
- Behm, J. K., Hagiwara, G., Lewiston, N. J., Quinton, P. M. & Wine, J. J. *Pediatr. Res.* **22**, 271–276 (1987).
- Sato, K. & Sato, F. *J. Lab. clin. Med.* **111**, 511–518 (1988).
- Horn, R. & Marty, A. *J. gen. Physiol.* **92**, 145–159 (1988).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85–100 (1981).
- Marty, A. & Neher, E. in *Single-Channel Recording* (eds Sakmann, B. & Neher, E.) 107–122 (Plenum Publishing Corporation, New York, 1983).
- McCann, J. D., Li, M. & Welsh, M. J. *J. gen. Physiol.* **94**, 1015–1036 (1989).
- Cliff, W. H. & Frizzell, R. A. *Ped. Pulm. Suppl.* **4**, 45 (1989).
- DiStefano, A. et al. *Pflügers Arch.* **405** (Suppl. 1), S95–S100 (1985).
- Boucher, R. C., Stutts, M. J., Knowles, M. R., Cantley, L. & Gatzky, J. T. *J. clin. Invest.* **78**, 1245–1252 (1986).
- Welsh, M. J. *J. Membrane Biol.* **88**, 149–163 (1985).
- Felgner, P. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7413–7417 (1987).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).

ACKNOWLEDGEMENTS. We thank B. Moss and T. Mizukami (NIH) for the plasmid pTM-1 and the recombinant vaccinia virus vTF7-3; A. Puga, P. Karp, E. Burton, C. Ward and C. Welsh for technical assistance; T. Mayhew for secretarial assistance; and J. Hunter and R. Fick for covering the MICU. This work would not have been possible without contributions of tissue by the CF patients, to whom we are especially indebted. This work was supported in part by the National Heart, Lung and Blood Institute and the National Cystic Fibrosis Foundation.

First simultaneous detection of PeV energy burst from the Crab Nebula

B. S. Acharya, M. V. S. Rao, K. Sivaprasad, B. V. Sreekantan & P. R. Vishwanath

Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Bombay 400 005, India

ALTHOUGH cosmic rays of energies up to 10^{20} eV have been detected, their origin remains obscure. Supernova remnants, pulsars and accreting binary systems containing neutron stars have been suggested as sources, but the link between these sites and cosmic rays cannot be directly established because interstellar magnetic fields render the distribution of charged particles spatially homogeneous. The discovery¹ of γ -rays of PeV (10^{15} eV) energy from Cygnus X-3, an X-ray binary, lends credibility to these sites as cosmic-ray sources, because γ -rays of such energy could only be produced by the interaction of very-high-energy particles with ambient matter. Systematic searches at TeV and PeV energies have revealed a few other possible sources, but the mostly sporadic nature of these detections has led to some doubt over their reality. Here we report the simultaneous detection on 23 February 1989, by experiments in the Kolar Gold Fields in southern India and the Baksan Valley of the northern Caucasus mountains in the USSR, of a PeV burst from the Crab Nebula, an object long suspected of being a cosmic-ray source.

Alexeeenko *et al.*² reported the observation of a PeV burst from the direction of the Crab Nebula during the period 14:00–19:00 UT on 23 February 1989 with their extensive air shower (EAS) array at Baksan (atmospheric depth, 840 g cm^{-2} ; longitude, 43°E ; latitude, 43°N). They recorded 57 events within 2.5° of the Crab direction, corresponding to an excess of 4.4σ over the expected background of 31.1 events.

Our EAS array is located at Kolar Gold Fields, India (KGF; atmospheric depth, 920 g cm^{-2} ; 78.3°E , 12.95°N). The meridian transit times of the Crab Nebula at KGF and Baksan on 23 February were 14:06 and 16:30 UT respectively. Because the atmospheric overburden increases with zenith angle, the primary energy required for a shower of a given threshold size (the total number of particles contained in the shower) to be recorded increases and thus the counting rate decreases with increasing zenith angle for a steep primary energy spectrum. In Fig. 1a, the zenith angle of the nebula on 23 February 1989 is plotted against time for both the sites. Although the most favourable zenith angles for the sites occur ~ 2.4 h apart, as expected from the difference in the meridian transit times, both the stations see the source at $\sim 25^\circ$ at $\sim 15:30$ UT. Thus, if the burst seen by the Baksan group is genuine, it should also be seen in the KGF data. We therefore searched the KGF data of 23 February 1989, starting at 13:15 UT, for any excess in the direction of the Crab Nebula.

The EAS array at KGF, described elsewhere^{3–6}, consists of 127 m^2 plastic scintillation detectors, 61 of which are equipped for nanosecond timing measurements, and 7 28.8-m^2 muon detectors ($>1 \text{ GeV}$), covering an effective area of $4.3 \times 10^4 \text{ m}^2$. The trigger rate is $\sim 1 \text{ Hz}$. A stable 5-MHz oscilloquartz crystal provides real-time information with an uncertainty of $\sim 1 \text{ ms}$. A software cut imposed during recording and analysis fixes the threshold shower size at $\sim 2 \times 10^4$ particles. For the present analysis only showers with cores within 100 m of the centre of the array were accepted. The arrival direction was obtained by fitting a conical shower front to the observed time delays taking into account the variation of shower disk thickness with core distance^{5,6}. The angular accuracy obtained by comparing the distribution of arrival directions of the same set of showers determined with odd and even numbered timing detectors

separately was 2.3° in space angle for showers of size $\geq 2 \times 10^4$ particles.

For observations of a given source, the zenith and azimuth angles of which change with time, the background is derived as follows. For an angular acceptance of θ_a around the source, we draw an annulus in the sky with rings around the zenith of radii $\theta = \theta_s - \theta_a$ and $\theta = \theta_s + \theta_a$, where θ_s is the zenith angle of the source at that time. Within this annulus, we accept showers that arrive with a space angle less than θ_a with respect to the source as 'source' events and all the rest 'background', which are then weighted according to the ratio of the two solid angles. As the contribution to the background comes from a larger solid angle, the accuracy of the background is very much higher. This procedure ensures that the zenith angle and the phase coverage are the same for the source and the background, so that gaps during observations, or short runs, and day-to-day variations in the number of events owing to malfunctioning of detectors do not matter, even when looking for long periodicities. This procedure requires uniform azimuthal angular distribution, which our data ensures⁴. We have verified that this distribution was uniform for the data of 23 February 1989. Here we use $\theta_a = 4^\circ$, which should contain $\sim 75\%$ of events from any point source.

The number of events recorded in 15-min intervals from the KGF and the Baksan arrays is plotted as a function of time in Fig. 1b. The data have been corrected for intervals in which there were gaps in recording (only three gaps of duration 5, 8 and 5 minutes in the entire run). The solid and dashed curves indicate the expected behaviour of the background taking into account the $\cos^2(\theta)$ angular distribution of cosmic-ray showers. It is clear that excess showers are seen from the direction of the Crab Nebula in successive periods from the two experiments, consistent with the variation of the source zenith angle at the two locations as a function of time. The excess is seen during the periods 13:15–16:00 UT in the KGF array and 15:00–19:00 UT in the Baksan array in the expected sequence, with peak counting rates corresponding to the meridian transit times at the two stations. Therefore, the burst lasted for several hours starting at 13:15 UT, or earlier and ending at 19:00 UT or later. In the KGF array 35 showers from the source were recorded

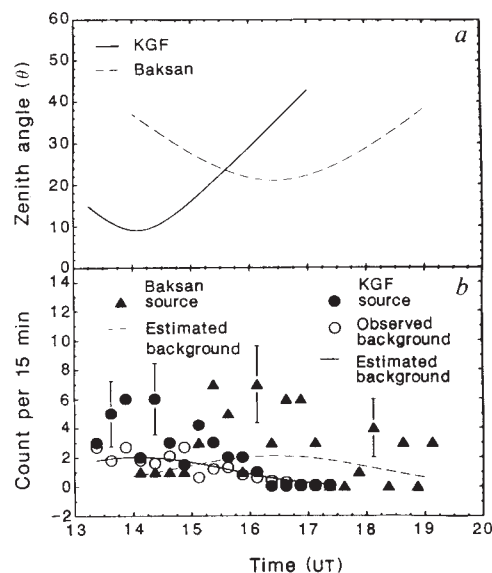


FIG. 1 a, Variation of the zenith angle of the Crab Nebula with time at KGF and Baksan on 23 February 1989. b, Counting rate in 15-min intervals of events arriving from the Crab direction plotted against time for both the KGF and Baksan data. The solid and dashed curves are the background rates of cosmic rays for KGF and Baksan arrays, respectively, estimated assuming an angular distribution of $\cos^2(\theta)$ and normalized to the total expected background over the respective periods. The open circles represent the background at KGF, calculated as described in the text.

during this period whereas the expected background was 17.8, which was based on 221 events. Using the likelihood ratio method of Li and Ma⁷, which also takes the uncertainty in the background estimate into account, the value of the parameter S was estimated as 3.4 and the corresponding probability P that the observed excess is purely due to chance is 3.4×10^{-4} . No other degree of freedom needs to be taken into account because only the data of 23 February were examined. To estimate the combined chance probability that the excess is seen on the same day in both the arrays, it is necessary to know the probability for the Baksan observation. Their results reported so far show that they looked for excesses on 270 days in 1985 and 63 days in 1989, but it is not clear whether they searched during other periods. When this becomes known, the combined probability would be smaller than 3.4×10^{-4} .

The muon content in the source and background events is determined in the following way. Showers from the source and the background regions are classified into various size groups and the muon information from all the seven detectors is classified into different core-distance groups. The average muon density and its error are determined for each of these groups. The non-zero densities in the distance range 30–150 m that are based on at least four observations are then normalized to a shower size of 10^5 and core distance of 50 m, using the experimentally obtained lateral distribution for a large sample of cosmic-ray showers, and their weighted mean and error estimated. The mean densities thus obtained for the source and background events respectively are 0.0397 ± 0.0016 and 0.041 ± 0.0005 . The ratio of the muon density in the excess showers from the Crab Nebula to that in background showers is estimated to be 0.93 ± 0.34 . Thus there is no clear evidence that the burst is due to high-energy γ -rays because in such showers the muon content is expected to be $\sim 10\%$ of the normal hadron showers. Based on this ratio we can say, at the 95% confidence level, that the muon content of the excess showers is $>37\%$ of that in normal cosmic-ray showers.

To see whether there was any signature of the Crab pulsar in the burst data, these events, recorded with millisecond uncertainty, were subjected to phase analysis by the epoch folding method using the pulsar elements $P = 33.36074816$ ms and $\dot{P} = 4.21072876 \times 10^{-13}$ appropriate to the epoch Julian date 47572.5 (15 February 1989), supplied by A. G. Lyne and R. S. Pritchard (personal communication). A barycentric time reduction program and related ephemeris (from D. Richards and I. Shapiro, respectively) were used to calculate the number of Crab pulsar cycles at the barycentre for every half hour at the observed site. The pulsar phase for each event was calculated by interpolation (accurate to better than 0.0016%) and folded to obtain the

phasogram for the pulsar, which is shown in Fig. 2. The 17 excess events seen in the burst (d.c. mode) are expected to be uniformly distributed, but the first half of the period, which includes the main pulse at 0 phase and the interpulse at 0.42 phase, contains all 17 excess events (26 from the source direction and 8.9 from the background). The Kolmogorov–Smirnov test⁸ for a uniform phase distribution gives a chance probability of 0.01 suggesting that the observed distribution is non-uniform and phase correlated.

Three possible explanations for the presence of large number of muons in the showers from Crab are: (1) there is a threshold energy above which photon interactions behave similar to hadron interactions^{9,10}; (2) the primaries are neutrinos that interact with a substantial cross-section at high energies according to some composite models¹¹; (3) a hitherto undiscovered, stable neutral particle is responsible for the showers. In the third case, because the signal is phase-locked for about half the pulsar period, the upper limit on the mass of the particle is $40 \text{ MeV } c^{-2}$ based on arguments similar to those in ref. 12.

With 17.2 ± 6.0 excess events recorded in 2.45 hours of real time over an area of $3.14 \times 10^8 \text{ cm}^2$ and with 60% efficiency for recording showers of energy $>10^{14} \text{ eV}$ (increasing to 100% for $E > 5 \times 10^{14} \text{ eV}$), the time-averaged flux of showers of energy $>10^{14} \text{ eV}$ arriving from the Crab Nebula during the KGF burst period is estimated to be $(1.3 \pm 0.4) \times 10^{-11} \text{ cm}^{-2} \text{ s}^{-1}$. The corresponding source luminosity, assuming an integral energy spectrum of E^{-1} with a high-energy cut-off at 10^{16} eV and isotropic emission, is $(4.8 \pm 1.6) \times 10^{36} \text{ erg s}^{-1}$. □

Received 26 February; accepted 7 September 1990.

- Samorski, M. & Stamm, W. *Astrophys. J.* **268**, L17–L20 (1983).
- Alexeenko, V. V. et al. *Proc. Int. Workshop on Very High Energy Gamma Ray Astronomy* (eds Stepanian, A. A., Fegan, D. J. & Cawley, M. F.) 187–191 (1989).
- Bhat, P. N. et al. *Techniques in Ultra High Energy Gamma Ray Astronomy* (eds Protheroe, R. J. & Stephens, S. A.) 1–7 (University of Adelaide, Adelaide, 1985).
- Bhat, P. N. et al. *Very High Energy Gamma Ray Astronomy* (ed. Turver, K. E.) 271–274 (Reidel, Dordrecht, 1987).
- Acharya, B. S. et al. *Cosmic Gamma Rays, Neutrinos and Related Astrophysics* (eds Shapiro, M. M. & Wefel, J. P.) 235–243 (Kluwer, Dordrecht, 1989).
- Sinha, S. et al. *Proc. 21 Int. Conf. Cosmic Rays Vol. 4* (ed. Protheroe, R. J.) 334–337 (University of Adelaide, Adelaide, 1990).
- Li, T. P. & Ma, Y. Q. *Astrophys. J.* **272**, 317–324 (1983).
- Frødesen, A. G., Skjæggstad, O. & Tofte, H. *Probability and Statistics in Particle Physics* (Universitetsforlaget, Bergen, 1979).
- Ochs, W. & Stodolsky, L. *Phys. Rev. D* **33**, 1247–1251 (1986).
- Drees, M. & Halzen, F. *Phys. Rev. Lett.* **61**, 275–277 (1988).
- Domokos, G. & Kovesi-Domokos, S. *Phys. Rev. D* **38**, 2833–2840 (1988).
- Yodh, G. B. *Physics and Astrophysics of Quark-Gluon Plasma* (eds Sinha, B. & Raha, S.) 230–246 (World Scientific, Singapore, 1988).

ACKNOWLEDGEMENTS. We thank the staff of the EAS project at KGF for their help and V. V. Alexeenko for information about the Baksan observations.

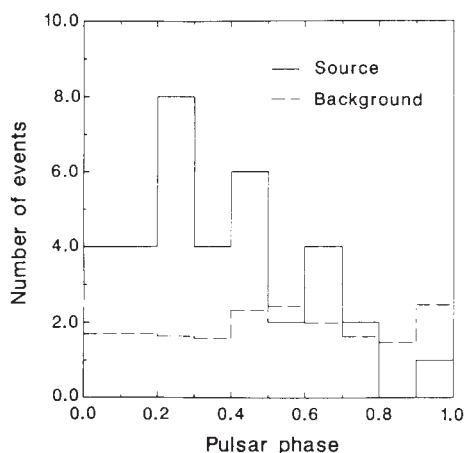


FIG. 2 The light curve for the showers arriving from the direction of the Crab pulsar between 13:15 and 16:00 UT on 23 February 1989, obtained using the ephemeris described in the text.

The 155-day solar period in the sixteenth century and later

S. M. Silverman

Physics Department, Boston College, Chestnut Hill, Massachusetts 02167, USA

RECENT interest in short solar periods (<1 yr) has been stimulated by the discovery of a period of ~ 150 – 160 days in very energetic solar flares¹ and hard-X-ray flares². The intensity of the peak seems to be very variable over extended intervals, and it is unclear whether it is a permanent feature of solar activity. Here I use auroral data to show the presence of this peak in data from 1570 to 1573 and in some other time periods, and to show its absence, or at least lack of prominence, at other times in the nineteenth and early twentieth centuries.

Magnetic and auroral data tend to lag the sunspot numbers and occasionally show very different behaviour (see, for example, ref. 3, where auroral activity at Jericho, Vermont is

TABLE 1 Presence or absence of the 150-day period

Period	No.*	Region	Range of sunspot number†	Presence of 150-day period
1570-72	27	Europe	(87-130) (M)	Yes (158)
1603-05	35	Europe	(70-88) (M)	No
1716-18	48	Europe	47-63 (M)	No
1736-39	117	Europe, Utrecht	70-111 (M)	Yes (154)
1748-51	78	Europe	47.7-83.4 (M)	No
1777-79	160	Europe	84.8-154.4 (M)	No
1787-89	228	Europe	118.1-132.0 (M)	No
1836-38	65	New England	103.2-138.3 (M)	No
1842-45	74	New England	10.7-40.1 (m)	Yes (150)
1847-50	120	New England	66.6-124.7 (M)	No
1854-56	79	New England	4.3-20.6 (m)	No
1858-61	203	New England	54.8-95.8 (M)	No
1864-67	193	New England	7.3-47.0 (m)	No
1870-72	301	New England	101.6-139.0 (M)	No
1876-79	91	New England	3.4-12.4 (m)	Yes (150, 141)
1892-94	405	United States	73.0-85.1 (M)	No
1900-02	126	United States	2.7-9.5 (m)	Yes (146)
1927-29	191	New England	64.9-77.8 (M)	No

Estimates of the sunspot numbers before 1700 are from ref. 22, estimates for dates after 1700 are from ref. 23.

* Number of days with aurora in the indicated time interval.

† M, sunspot maximum occurred during the time interval; m, sunspot minimum occurred during the time interval.

compared with both magnetic activity and sunspot numbers for the period 1883-1931 and ref. 4 in which auroral activity is compared with both magnetic activity and sunspot number over a more extended time interval). These parameters, however, seem to be related to an underlying solar activity⁵. Despite these differences, auroral and magnetic data, used and interpreted cautiously, may be used as proxy data for solar activity. For example, a 1.4-year period found in auroral activity⁶ is consistent with a similar period found in sunspot activity and geomagnetic data^{7,8} and in flare occurrence⁹. Thus, where the results of analysis of auroral and magnetic data are consistent with those of solar data one may safely assume that the activity is related to solar activity.

The 155-day period (although values between 150 and 160 days have been reported, I refer to it here as the 155-day period, for convenience) has now been found in a number of solar features. (For a summary of periods in the range 100-500 days, see ref. 10.) Almost all of these studies have involved only solar cycles 19, 20 and 21, although a peak corresponding to the 155-day period has been reported by Wolf¹¹ based on Zurich sunspot numbers from 1874-1979. The period has been reported as stronger in cycle 21 than in cycle 20. In cycle 19, the 155-day period is found in flares in the Sun's southern hemisphere, but it is of low statistical significance; it is not prominent or present in northern hemisphere or whole-disk flares¹². The peak has also been reported in measurements of the solar diameter during the latter part of the Maunder Minimum, 1683-1718^{13,14}.

The auroral data for New England that I use here are drawn from a database derived from manuscript material in the US National Archives (S.M.S., unpublished) and from published data. US data were taken from publications (S.M.S., unpublished). Data from Utrecht, the Netherlands for 1736-39 are taken from a series of observations by Musschenbroek that were abstracted from his manuscripts by Lovering¹⁵. These data were then included in the catalogue of Fritz¹⁶ (1873) and recently, with additions, in a catalogue of auroral dates from 1000-1900 by Krivsky and Pejml¹⁷ (1988). For dates before the nineteenth century I have relied on the Krivsky and Pejml catalogue, which is drawn from a variety of mostly published sources. I have denoted the data sets taken from this catalogue as 'Europe', because most of the data included are from Europe although some data are from the Orient. All data in the Krivsky and Pejml catalogue are for latitudes <55°.

Blackman-Tukey power spectra were used for the analyses of periodicities^{6,18}. For the auroral data, days on which an aurora

occurred were given a value of 1, and a value of 0 was assigned to days when no aurora was noted. For the New England data before 1890 the occurrence of aurora was accepted when at least two stations reported aurora. For later New England data, where the published reports often merely state 'New England', every date was used.

Because the original reports of the 155-day period emphasized energetic flares as the source, and as flare occurrence may be expected to be closely correlated with high sunspot numbers, three- or four-year periods around sunspot maxima were initially chosen for analysis. For New England and the United States, periods from 1836 to 1929 were used. Periods earlier than this were chosen where it seemed that sufficient data existed to provide a reasonable sample. Periods around sunspot minima were later added. For New England and the United States these ranged from 1842 to 1902. For earlier periods the data were considered to be too sparse to provide reliable analyses.

For the years 1570-73 a peak in the data from Europe is found between about 158 and 165 days. The resolution of these spectra is not sufficient to determine the position with greater precision. The sample 1570-73 gives a peak at ~165 days, whereas the sample for 1570-72 gives a peak at ~158 days. As the database is relatively small it is not clear whether this difference is due to the database or to a difference in the time periods used. In any case, it seems clear that a peak in the vicinity of 160 days is present. Figure 1a shows the portion of the spectrum with periods greater than 86 days. The only prominent peak in the region below 230 days is that at ~158 days.

The results for other samples are shown in Table 1. In addition to these, no peak is found in the spectrum of the daily sunspot number for 1859-61, which tends to confirm the absence of the peak in the New England auroral data. Two spectra were run

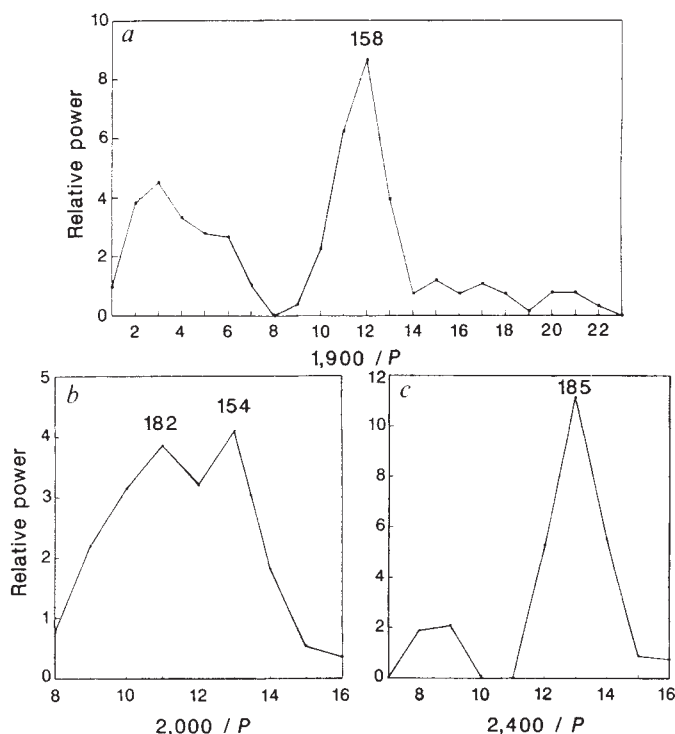


FIG. 1 a, Power spectrum of auroras, 1570-72, for periods (p) greater than 86 days. The main peak is at 158 days. The two points at the neighbouring lags correspond to periods of 146 and 172 days. b, Power spectrum of auroras, 1736-39 for periods from 125 to 250 days. The two peaks are at about 182 and 154 days. The peak at ~182 days may be the semi-annual peak of 180 days. c, Power spectrum of auroras, 1787-90 for periods from 133 to 343 days. The large peak is at ~185 days, and may be the semi-annual peak. No peak is evident at ~160 days.

of the daily magnetic activity index, *aa*. In one, for 1870–72, no peak was found. In the other, for 1876–78, a peak at 164 days was found, which may be compared with the peak at ~150 days found in the New England data for 1877–79. A spectrum of the number of magnetic storms with *aa* > 60 in 1892–94 shows no peak in the region of interest.

Figure 1b, c show the spectra in the vicinity of 160 days for the periods 1736–39 and 1787–90. The spectrum from the earlier period shows a peak at ~154 days as well as one at ~182 days. For the later interval, the 150-day peak is either absent or is present only as a weak shoulder on the strong 180-day peak. The relative amplitudes of the approximately semi-annual peak (180 days) are consistent with Swedish data, over the period 1721–1943⁶ (for recent discussions of the annual variation, see refs 19, 20).

These results indicate that the 155-day periodicity is only infrequently prominent. The presence of a peak near this period in data going back to the sixteenth century, however, demonstrates that it is a persistent feature of solar activity over intervals of centuries, and confirms the report of the presence of the peak during the Maunder Minimum¹³. The finding that the period is sporadic is consistent with the results of the past three solar cycles, in which the peak is not observed clearly in whole-disk behaviour for the nineteenth cycle (see Fig. 1 of ref. 12), and in which it is weaker in the twentieth than in the twenty-first cycle¹⁰. The reality of this erratic peak seems to be established, both by the criteria used in the analyses and by its occurrence in a variety of observational parameters. A final comment may be made on the position of the ~150-day peak. I have assumed

that peaks between 146 and 160 days are related because of resolution problems in data and the method of analysis. Table 1 indicates a slow drift over the centuries in the position of this peak. The occurrence of such fuzzy periodicities is not unusual or unexpected (see, for example, ref. 21). If real, this drift may reflect a basic change in a so-far unknown solar property. □

Received 11 December 1989; accepted 7 August 1990.

1. Rieger, E. *et al.* *Nature* **312**, 623–625 (1984).
2. Kiplinger, A. L., Dennis, B. R. & Orwig, L. E. *Bull. Am. astr. Soc.* **16**, 891 (1984).
3. Silverman, S. M. & Blanchard, D. *Planet. Space Sci.* **31**, 1131–1135 (1983).
4. Legrand, J.-P. & Simon, P. *Annls Geophys.* **45**, 161–168 (1987).
5. Feynman, J. & Gu, X. Y. *Rev. Geophys.* **24**, 650–666 (1986).
6. Silverman, S. M. & Shapiro, R. *J. Geophys. Res.* **88**, 6310–6316 (1983).
7. Akioka, M., Kubota, J., Suzuki, M., Ichimoto, K. & Tohmura, I. *Solar Phys.* **112**, 313–316 (1987).
8. Yacob, A. & Bhargava, B. N. *J. atmos. terr. Phys.* **30**, 1907–1911 (1968).
9. Ichimoto, K., Kubota, J., Suzuki, M., Tohmura, I. & Kurokawa, H. *Nature* **316**, 422–424 (1985).
10. Lean, J. L. & Brueckner, G. E. *Astrophys. J.* **337**, 568–578 (1989).
11. Wolff, C. L. *Astrophys. J.* **264**, 667–676 (1983).
12. Bai, T. *Astrophys. J.* **318**, L85–L91 (1987).
13. Ribes, E., Merlin, Ph., Ribes, J.-C. & Bartholot, R. *Annls Geophys.* **7**, 321–330 (1989).
14. Ribes, E. *et al.* in *The Sun in Time* (eds Sonett, C. P., Giampapa, M. S. & Matthews, M. S.) (University of Arizona Press, Tucson, in the press).
15. Lovering, J. *Mem. Am. Acad. Arts Sci.* **10**, 9–351 (1866–1871).
16. Fritz, H. *Verzeichnis Beobachteter Polarlichter* (Wien, 1873).
17. Krivsky, L. & Pejml, K. *Publs astr. Inst. Czech. Acad. Sci.* **75** (1988).
18. Blackman, R. B. & Tukey, J. W. *The Measurement of Power Spectra* (Dover, New York, 1959).
19. Russell, C. T. *Geophys. Res. Lett.* **16**, 555–558 (1989).
20. Silverman, S. M. in *Solar Wind-Magnetosphere Coupling* (eds Kamide, Y. & Slavin, J. A.) 643–654 (Terra Scientific, Tokyo, 1986).
21. Silverman, S. M. in *Proc. 11th Conf. on Probability and Statistics in Atmospheric Sciences* 301–304 (Am. met. Soc., Boston, 1989).
22. Schöve, D. J. (ed.) *Sunspot Cycles*, 10 (Hutchinson & Ross, Stroudsburg, 1983).
23. McKinnon, J. A. *Sunspot Numbers. 1610–1985, Rep. UAG-95* (World Data Center, Boulder, 1987).

ACKNOWLEDGEMENTS. This work was supported by NASA.

The energy spectrum of knots and links

H. K. Moffatt

Department of Applied Mathematics and Theoretical Physics,
Silver Street, Cambridge CB3 9EW, UK

KNOTTED and linked structures arise in such disparate fields as plasma physics, polymer physics, molecular biology and cosmic string theory. It is important to be able to characterize and classify such structures. Early attempts to do so¹ were stimulated by Kelvin's² recognition of the invariance of knotted and linked vortex tubes in fluid flow governed by the classical Euler equations of motion. The techniques of fluid mechanics are still very natural for the investigation of certain problems that are essentially topological in character. Here I use these techniques to establish the existence of a new type of topological invariant for knots and links. Any knot or link may be characterized by an 'energy spectrum'—a set of positive real numbers determined solely by its topology. The lowest energy provides a possible measure of knot or link complexity.

Continuous deformation of a knotted structure may be achieved by embedding the structure in an incompressible fluid medium moving with continuous velocity $\mathbf{v}(\mathbf{x}, t)$ (where $\nabla \cdot \mathbf{v} = 0$), convecting and distorting the structure in the process. This flow induces a continuous, time-dependent, orientable and volume-preserving mapping $\mathbf{x} \rightarrow \mathbf{X}(\mathbf{x}, t)$ of the medium onto itself, where $\mathbf{X}(\mathbf{x}, t)$ represents the position at time t of the fluid particle that passes through position $\mathbf{x}$ at time $t = 0$.

Let $\mathbf{B}(\mathbf{x}, t)$ be any solenoidal vector field ($\nabla \cdot \mathbf{B} = 0$) convecting with the fluid under the condition that the flux of $\mathbf{B}$ through any closed circuit moving with the fluid is conserved; such a field may be described as 'frozen' in the fluid, and it satisfies the frozen-field equation³

$$\partial \mathbf{B} / \partial t = \nabla \times (\mathbf{v} \times \mathbf{B}) \quad (1)$$

The solution can be obtained in the form

$$\mathbf{B}_t(\mathbf{X}, t) = \mathbf{B}_j(\mathbf{x}, 0) \partial \mathbf{X}_i / \partial \mathbf{x}_j \quad (2)$$

a result that encapsulates both the convection of the field from position $\mathbf{x}$ to $\mathbf{X}$ and its simultaneous rotation and distortion by the deformation tensor $\partial \mathbf{X}_i / \partial \mathbf{x}_j$. Equation (2) establishes a topological equivalence between the initial field $\mathbf{B}_0(\mathbf{x}) = \mathbf{B}(\mathbf{x}, 0)$ and the field $\mathbf{B}(\mathbf{X}, t)$; in particular, all knots and links in the field-line structure are conserved under the deformation described by equation (2).

To exploit the properties of equations (1) and (2) in the search for topological invariants of a knot K , it is first necessary to construct a tubular neighbourhood $\mathcal{T}_K$ of the knot, and a knot field $\mathbf{B}_K(\mathbf{x})$ confined to this neighbourhood, each field line being a satellite of K . Care is needed, however, in considering the twist of the field in $\mathcal{T}_K$.

Consider a knot K of length L_0 in a configuration for which a plane projection exhibits a minimum number of crossings. The knot may be deformed inextendibly to lie entirely in the plane $z = 0$ except for smooth indentations into the half-space $z < 0$ at each of the underpasses. By a finite number of reflections of underpasses in the plane $z = 0$, a related unknotted curve C may be formed, which may be continuously deformed to a circle C_0 of radius $R = L_0/2\pi$.

Conversely, one may start with a circle, C_0 , $x^2 + y^2 = R^2$, and define a tubular neighbourhood $\mathcal{T}_0$ of C_0 , in cylindrical polar coordinates (r, φ, z) , by

$$(r - R)^2 + z^2 < (\varepsilon R)^2 \quad (3)$$

where $\varepsilon < 1$; the limiting situation when $\varepsilon \rightarrow 0$ is of particular interest. The area of cross-section of $\mathcal{T}_0$ is $A_0 = \pi \varepsilon^2 R^2$, and its volume is $V = 2\pi^2 \varepsilon^2 R^3$. This tube can be made into a flux tube by defining a field $\mathbf{B}_0(\mathbf{x})$ that is zero outside $\mathcal{T}_0$, and which has the form

$$\mathbf{B}_0(\mathbf{x}) = (0, 2\pi r \Phi / V, 0) \quad (4)$$

inside $\mathcal{T}_0$, where Φ is constant (the flux of $\mathbf{B}_0(\mathbf{x})$ across any meridian section of the tube). This field is invariant under any

axisymmetric volume-preserving rearrangement of field lines; this follows from equation (1), which for axisymmetric convection of a toroidal field becomes $D(B_\phi/r)/Dt=0$, where D/Dt is the Lagrangian derivative ($=\partial/\partial t + \mathbf{v} \cdot \nabla$).

The knot field $\mathbf{B}_K(\mathbf{x})$ is obtained in three steps. (1) The flux tube is cut at any section $\varphi = \text{constant}$, twisted through angle $2\pi h_0$, the twist being uniformly distributed with respect to the angle φ , and reconnected (Fig. 1). If $h_0 = +1$, then every pair of $\mathbf{B}$ lines is now simply linked, and the associated helicity^{4,5} of the field, is obtained by integrating over the flux elements $d\Phi_1$

$$\mathcal{H}_0 = 2 \int_0^\Phi \Phi_1 d\Phi_1 = \Phi^2 \quad (5)$$

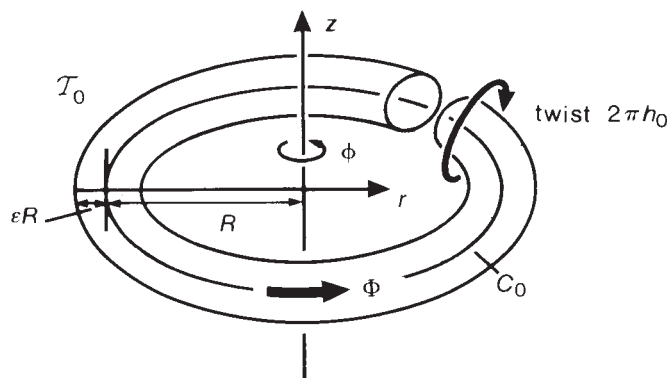


FIG. 1 The flux tube $\mathcal{T}_0$ (equation (3)) is a torus of circular section centred on the circle C_0 , and carrying flux Φ . Twist is introduced by cutting the tube at a section $\varphi = \text{constant}$, twisting through an angle $2\pi h_0$, and reconnecting. The helicity thus generated is $h_0\Phi^2$.

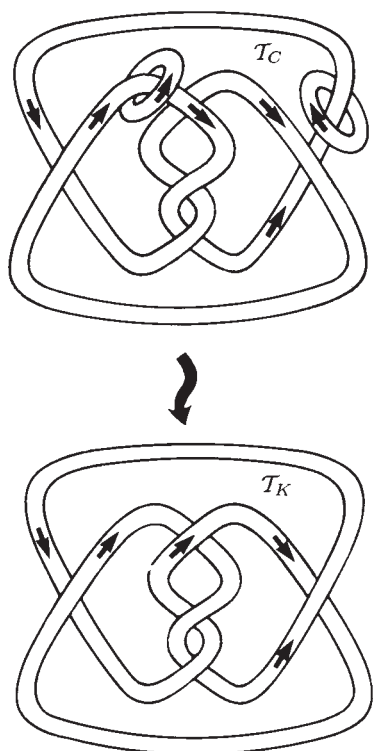


FIG. 2 The unknotted tube $\mathcal{T}_C$ is converted to the knotted tube $\mathcal{T}_K$ by switching a number of crossings. In the example shown here, two positive crossings are switched to become negative crossings; each switch is equivalent to the insertion of a small loop that cancels the field on one side and makes it reappear on the other. The change in helicity associated with the switches in this case is $-4\Phi^2$, so that $h = h_0 - 4$.

More generally, the helicity is

$$\mathcal{H}_0 = h_0\Phi^2 \quad (6)$$

(2) The tube is now deformed by a continuous incompressible flow field $\mathbf{v}_1(\mathbf{x}, t)$ ($0 \leq t \leq T$), with associated mapping $\mathbf{x} \rightarrow \mathbf{X}_1(\mathbf{x}, T)$, so that $C_0 \rightarrow C$ and $\mathcal{T}_0 \rightarrow \mathcal{T}_C$, a tubular neighbourhood of the curve C defined above; both Φ and V , and also the helicity $\mathcal{H}_0$ are conserved in this process. Here, ε may be assumed sufficiently small that $\mathcal{T}_C$ does not intersect itself. (3) The field is reflected in the relevant indentations about $z=0$ so that the flux tube $\mathcal{T}_C$ becomes a flux tube $\mathcal{T}_K$ containing the knot K (Fig. 2). Again, Φ and V are conserved. The example of Fig. 2 shows that the helicity changes by $\pm 2\Phi^2$ for every reflection that converts a negative (positive) crossing to a positive (negative) one; hence $\mathcal{H}_0 \rightarrow \mathcal{H} = h\Phi^2$ where

$$h = h_0 + 2(N_+ - N_-) \quad (7)$$

where N_+ and N_- are the numbers of such crossings required to yield the knot K . There may well be different ways of achieving this end (different sets of reflections may yield the same knot) but because h_0 is arbitrary, any desired value of h , positive or negative or zero, may be achieved, and this freedom corresponds to the freedom in specifying the twist of the field in the flux tube $\mathcal{T}_K$.

The choice $h_0 = -2(N_+ - N_-)$ makes $h = 0$, that is, the field helicity is zero; this corresponds to 'zero-framing' of the knot as discussed in a physical context by Witten⁶; I adopt this natural choice here. The topology of the field $\mathbf{B}_K(\mathbf{x})$ thus constructed is then uniquely determined.

The knot field is allowed to 'relax' by a procedure that has proved effective in considering the existence and stability of magnetostatic equilibria⁷. For any field $\mathbf{B}(\mathbf{x}, t)$ that is non-zero only in some bounded volume, the field energy may be defined as

$$M(t) = \frac{1}{2} \int \mathbf{B}^2 dV \quad (8)$$

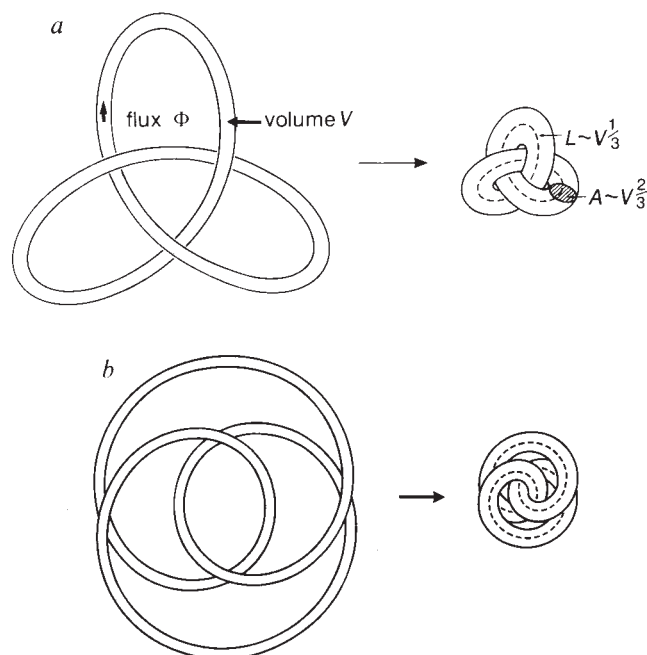


FIG. 3 Relaxation of a flux tube knotted in the form of a trefoil knot, with V and Φ conserved; relaxation is arrested when the axial length L is of the order of $V^{1/3}$ and the mean cross-sectional area A is of the order of $V^{2/3}$. The energy is then $M^E = m\Phi^2 V^{1/3}$, where m is a dimensionless positive real number, a topological invariant of the knot. a), The presentation $K_{2,3}$ of the trefoil for which $m = m_0$; b), the presentation $K_{3,2}$ for which I conjecture that $m = m_1 > m_0$.

which, by virtue of equation (1), satisfies

$$\frac{dM}{dt} = \int \mathbf{B} \cdot [\nabla \times (\mathbf{v} \times \mathbf{B})] dV = - \int \mathbf{v} \cdot [(\nabla \times \mathbf{B}) \times \mathbf{B}] dV \quad (9)$$

I use the initial condition

$$\mathbf{B}(\mathbf{x}, 0) = \mathbf{B}_K(\mathbf{x}) \quad (10)$$

and assume that $\mathbf{v}(\mathbf{x}, t)$ is itself instantaneously determined by the equation

$$k\mathbf{v} = -\nabla p + (\nabla \times \mathbf{B}) \times \mathbf{B} \quad (11)$$

where k is a positive constant and p is a 'pressure' field determined by the condition that $\nabla \cdot \mathbf{v} = 0$ for all t . Both $\mathbf{v}$ and ∇p are at most of the order of $|\mathbf{x}|^{-3}$ as $|\mathbf{x}| \rightarrow \infty$. Combining equations (9) and (11)

$$\frac{dM}{dt} = -k \int \mathbf{v}^2 dV \quad (12)$$

so that $M(t)$ is indeed monotonic decreasing if $\mathbf{v}$ is not identically zero.

It has been shown⁸ that if the topology of $\mathbf{B}(\mathbf{x}, 0)$ is non-trivial, then the energy of any field obtainable by a transformation of the form of equation (2) has a positive lower bound. Hence, if the knot K is non-trivial, the energy function $M(t)$ is bounded below and, by virtue of equation (12), must tend to a limit $M^E (> 0)$, as $t \rightarrow \infty$. Here the superscript E indicates that this is an equilibrium state for which $\mathbf{v} = 0$.

Consider how the decrease of energy described by equation (12) actually occurs. At the initial instant

$$L_0 = 2\pi(V/\pi\epsilon^2)^{1/3} \quad \text{and} \quad A_0 = V/L_0 \quad (13)$$

and, if it is assumed that $\epsilon \ll 1$, then

$$M(0) \approx \frac{1}{2}(\Phi/A_0)^2 V = 8\Phi^2 V^{-1/3}(\pi/\epsilon)^{1/3} \quad (14)$$

As Φ and V are invariant during the relaxation process, the energy can decrease only through increase of the average cross-sectional area A of $\mathcal{T}_K$ and consequent decrease of (axial) tube length L . This process is illustrated for the trefoil knot in Fig. 3. It is evident that the process of energy reduction must come to a halt when different parts of the flux tube come in contact with each other. It is this topological barrier that implies the positive lower bound for $M(t)$ referred to above. The lower bound is attained when L and A are determined by V alone (in conjunction with the knot topology) that is, $L \sim V^{1/3}$, $A \sim V^{2/3}$. Then on dimensional grounds (compare to equation (14))

$$M^E = m\Phi^2 V^{-1/3} \quad (15)$$

where m is a positive real number, determined only by the knot topology. By virtue of its construction, m is a topological invariant of the knot.

There is no guarantee that the end state as $kt \rightarrow \infty$ is uniquely determined, and indeed it seems likely that for knots of any complexity, a variety of distinct asymptotic configurations may be attainable starting from different initial geometrical configurations of the same knot. In this case, different values of m will in general be attained for different end states. I denote these by m_i ($i = 0, 1, 2, \dots$) ordered so that $0 < m_0 \leq m_1 \leq m_2 \leq \dots$. The sequence $\{m_0, m_1, m_2, \dots\}$ may then be described as the energy spectrum of the knot, with m_0 the ground state energy. Generally speaking, a high value of m_0 will indicate a complex knot and indeed m_0 could reasonably be used as a measure of knot complexity.

The suggestion that there is in general a multiplicity of possible end states derives support from consideration of the two different forms of the trefoil knot (usually denoted $K_{2,3}$ and $K_{3,2}$) shown in Fig. 3. I conjecture that these may relax to distinct stable end states with distinct energy levels m_0 (for $K_{2,3}$) and m_1 (for $K_{3,2}$).

Very similar considerations apply to the problem of characterizing links of two or more components. Considering the case of two linked curves C_1 and C_2 , flux tubes of equal volume V

and carrying equal flux Φ around C_1 and C_2 can be constructed by the same procedure as for knots, and thus an initial link field $\mathbf{B}_L(\mathbf{x})$ can be generated for the relaxation problem. Again the asymptotic energy has the form of equation (15), where the real number m is a topological invariant for the link in question.

In the four physical contexts mentioned at the outset, the energy of knotted and linked structures has a central role in formulating a self-consistent theory. In polymer physics, for example, a tangled macromolecule will tend to relax towards the 'ground-state' configuration compatible with its topology. The result that the energy of a structure, if suitably defined, has a minimum value that is unambiguously related to its topology, seems to be new, and may suggest techniques whereby topologically distinct structures may be detected. Related concepts are discussed in recent papers by Freedman and He^{9,10}. $\square$

Received 26 April; accepted 6 August 1990.

1. Tait, P. G. *Scientific Papers* Vol. 1, 273–347 (Cambridge University Press, 1898).
2. Kelvin, Lord (then W. Thomson) *Trans. R. Soc. Edinb.* **25**, 217–260 (1868).
3. Batchelor, G. K. *Proc. R. Soc. A* **213**, 349–366 (1952).
4. Moffatt, H. K. *J. Fluid Mech.* **35**, 117–129 (1969).
5. Berger, M. A. & Field, G. B. *J. Fluid Mech.* **147**, 133–148 (1984).
6. Witten, E. in *Braid Group, Knot Theory and Statistical Mechanics* (ed. Young, C. N. & Ge, M. L.) 239–329 (World Scientific, Singapore, 1989).
7. Moffatt, H. K. *J. Fluid Mech.* **159**, 359–378 (1985).
8. Freedman, M. H. *J. Fluid Mech.* **194**, 549–551 (1988).
9. Freedman, M. H. & He, Z.-X. *Topology* (in the press).
10. Freedman, M. H. & He, Z.-X. Preprint, Res. Rep. GCG 15, Geometry Supercomputers Project, University of Minnesota (1990).

ACKNOWLEDGEMENTS. I thank R. Lickorish who pointed out an error in an earlier draft of this note.

Tunnelling evidence for predominantly electron–phonon coupling in superconducting $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ and $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$

Q. Huang*, J. F. Zasadzinski†, N. Tralshawala†, K. E. Gray*, D. G. Hinks*, J. L. Peng‡ & R. L. Greene‡

* Materials Science Division, and Science and Technology Center for Superconductivity, Argonne National Laboratory, Argonne, Illinois 60439, USA

† Illinois Institute of Technology, Chicago, Illinois 60616, USA

‡ Center for Superconductivity Research, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742, USA

AMONG the superconducting oxides, the cubic, copper-free $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ (BKBO) and the electron-doped compound $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ (NCCO) stand out as being somewhat different from the rest. Nevertheless, an understanding of the pairing mechanism in BKBO (transition temperature, $T_c \approx 30$ K) and NCCO ($T_c \approx 23$ K) may give important insights into the mechanisms of the higher- T_c superconductors. Here we report tunnelling spectroscopy measurements on BKBO and NCCO, using point-contact junctions that exhibit low leakage currents and sharp conductance peaks at the gap voltages $V = \pm \Delta/e$. Reasonably symmetric and reproducible structures are observed in the high-bias tunnelling conductances which are characteristic of phonon effects as seen in conventional superconductors. We have inverted the tunnelling data and obtained the Eliashberg functions, $\alpha^2 F(\omega)$, where $F(\omega)$ is the phonon density of states at energy $\hbar\omega$. For BKBO, $\alpha^2 F(\omega)$ bears a close resemblance to the available phonon density of state determined by inelastic neutron scattering, most importantly consistently reproducing the minima. The fact that the $\alpha^2 F(\omega)$ are not identical for different junctions leads to some uncertainty, but the calculated values of T_c are in good agreement with experiment for both BKBO and NCCO. Also, there is a good match between the calculated values of the total

TABLE 1 Measured and calculated parameters

Junction	Δ (meV)	λ	μ^*	T_c (K)	T_c^{calc} (K)
BKBO-1	3.8	1.2 ± 0.2	0.11 ± 0.04	24.5 ± 3	19
BKBO-2	4.5	1.3 ± 0.3	0.04 ± 0.04	24.5 ± 3	20
NCCO-1	3.7	0.9 ± 0.1	0.05 ± 0.05	22.0 ± 2	21
NCCO-2	3.7	1.0 ± 0.1	0.08 ± 0.04	22.0 ± 2	19

Magnetic measurements determined T_c at the mid-point (50%) and the transition widths are the 10 and 90% values. The errors on the calculated values of λ and μ^* are estimates based on the inversion analysis.

electron-phonon coupling constant λ and the measured $2\Delta/kT_c$, consistent with predominantly phonon-mediated pairing mechanisms in these compounds.

The electron-phonon interaction in conventional superconductors is determined from measurements of the normalized tunnelling conductance, σ_s/σ_n , where $\sigma = dI/dV$, I is the current, V is the voltage and s and n are the superconducting and normal states, respectively¹. At low temperatures, σ_s/σ_n is just the strong-coupling analogue, $N(E)$, of the BCS density of states, $N_{\text{BCS}}(E)$. Thus

$$\sigma_s/\sigma_n = N(E) = \text{Re} \{E/[(E^2 - \Delta^2(E))]^{1/2}\} \quad (1)$$

where $E = eV$, e is the electronic charge and $\Delta(E)$ is the complex, energy-dependent gap function. At low energies, $\Delta(E)$ is the ordinary BCS gap parameter Δ but near higher-energy peaks in $F(\omega)$, it has a strong energy-dependence and a large imaginary part related to the electron lifetime. Phonons decrease the lifetime, such that σ_s/σ_n deviates from $N_{\text{BCS}}(E)$ (a constant Δ in equation (1)), in general by a few per cent or less in the region of phonon peaks. This deviation scales with the overall electron-phonon coupling strength and, for example, in Nb a deviation

of $\sim 1\%$ near the phonon energies of 15 and 25 meV gives $\lambda \approx 1$ (ref. 2).

The electron-phonon spectral function (or Eliashberg function), $\alpha^2 F(\omega)$, is the principal quantity of interest, where α^2 is a measure of the coupling strength of electrons to phonons of energy $\hbar\omega$. The McMillan-Rowell (MR) procedure³ for extracting $\alpha^2 F(\omega)$ from tunnelling data uses an iterative approach to solve the Eliashberg equations. For the procedure to be meaningful, the MR iteration must converge with the resulting Coulomb repulsion term, μ^* , exhibiting a reasonable value and $\alpha^2 F(\omega)$ replicating distinctive features (peak and especially valley energies) of the phonon density of states. A more stringent test is to demonstrate that these results quantitatively reproduce the measured voltage-dependence of σ_s/σ_n and predict a T_c close to the experimental one. Earlier studies⁴ of thin-film junctions on BKBO gave evidence for phonon structures in the voltage-dependence of σ_s/σ_n , but no quantitative results were obtained.

Polycrystalline samples of $\text{Nd}_{1.85}\text{Ce}_{0.15}\text{CuO}_{4-y}$ were prepared by a solid-state reaction from high-purity Nd_2O_3 , CeO_2 and CuO (ref. 5). Stoichiometric mixtures were thoroughly ground and heated in air using an Al_2O_3 crucible at $1,000^\circ\text{C}$ for 12 h. The resulting powders were reground, pressed into pellets and refired in air at $1,100^\circ\text{C}$ for 12 h, followed by quenching to room temperature. To obtain superconductivity, it was necessary to anneal the sintered pellets in a reducing atmosphere. High-density (near 100% of the theoretical value) samples of $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ (nominally with $x = 0.375$) were prepared from the oxide powders (Bi_2O_3 , BaO and KO_2) in a melt-process technique described previously⁶. The Meissner fraction was $>50\%$ for both materials, indicating bulk superconductivity. The mid-point and transition widths of field-cooled and zero-field-cooled magnetization measurements as a function of temperature are given in Table 1.

Tunnelling spectra were obtained with a point-contact apparatus recently used for measurements of the high- T_c oxides, $\text{Bi}_{1.85}\text{Pb}_{0.15}\text{Sr}_2\text{CaCu}_2\text{O}_y$ and $\text{Ti}_2\text{Ba}_2\text{CaCu}_2\text{O}_y$ ^{7,8}, and for high-resolution tunnelling spectroscopy on Nb single crystals⁹. The latter showed that this method can reproduce the precision obtained with thin-film junctions in measuring the small con-

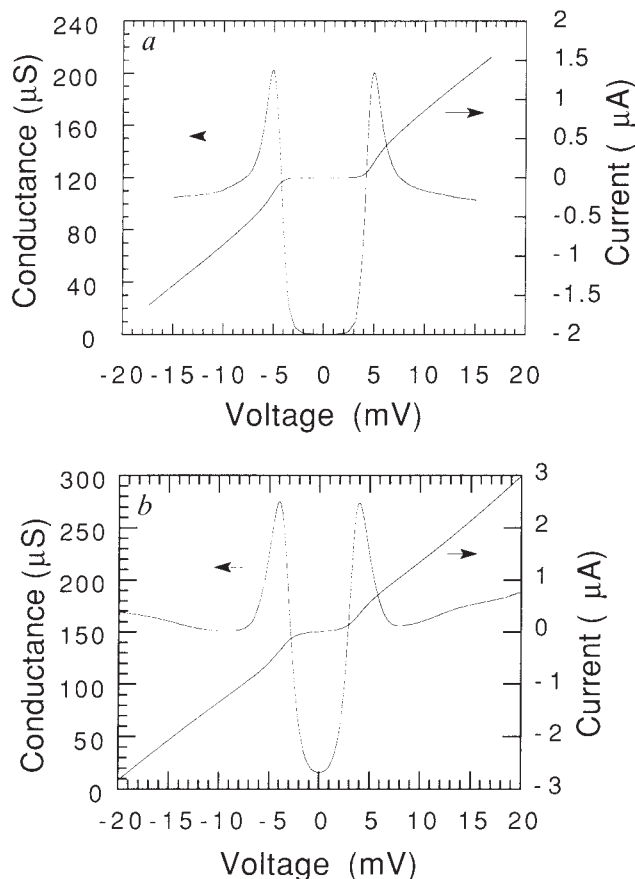


FIG. 1 Measured I - V and σ_s - V for point-contact tunnel junctions on a, BKBO and b, NCCO.

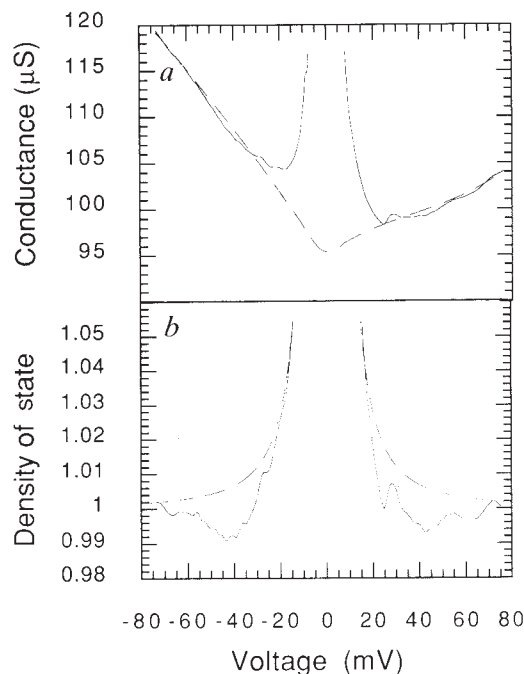


FIG. 2 a, Expanded scale of σ_s (solid line) and σ_n (dashed line) in the region of phonon structure and b, σ_s/σ_n (solid line) and $N_{\text{BCS}}(E)$ (dashed line) for a BKBO tunnel junction.

ductance changes ($\sim 1\%$) that are the signature of phonon effects. All measurements were done using a gold tip as the counter-electrode. The current-voltage characteristics and $\sigma = dI/dV$ in the gap region are shown in Fig. 1 at $T = 4.2$ K for BKBO and NCCO. The BKBO data of Fig. 1a exhibit nearly ideal junction characteristics, with $\sigma_s/\sigma_n(V=0) < 1\%$, and can be fit by equation (1) with only thermal smearing. In the case of NCCO (Fig. 1b), the $\sigma_s/\sigma_n(V=0)$ of $\sim 10\%$ is significantly smaller than previous studies of this compound¹⁰, which went up to $\sim 50\%$, but larger than expected for just thermal smearing in an ideal junction. It is not clear whether this is due to leakage currents or the general broadening commonly observed in tunnelling in high- T_c superconductors.

For NCCO, an energy gap of $\Delta = 3.7 \pm 0.1$ meV was found in all 15 junctions measured: the ratio $2\Delta/kT_c = 3.9 \pm 0.4$ implies moderate coupling strength. These are in much better agreement with infrared reflectance measurements¹¹ than previous tunnelling conductance data¹⁰, which were significantly broadened in the gap region. The gap values for BKBO varied from point-to-point on the polycrystalline pellet, ranging from 3.8 to 4.6 meV, also in excellent agreement with infrared data¹². For each point on the pellet, the σ_s/σ_n curves fit very well with thermal smearing only: the absence of further smearing indicates that tunnelling was usually into one grain only (the grain diameters are estimated from electron diffraction studies to be greater than $1 \mu\text{m}$, and much larger than the Au-tip radius). The different gap values are probably a consequence of the Au tip contacting grains having different K composition and therefore different Δ and T_c . If we associate the largest (smallest) gap with the 10% (90%) point of the magnetic transition, we find the same moderate coupling strength for BKBO ($2\Delta/kT_c = 3.8$ – 3.9) as for NCCO.

An example of the phonon structures in the tunnelling conductance is shown in Fig. 2a, in which symmetric structures near ± 25 , ± 45 and ± 60 mV are observable in the superconducting data (solid line), but not in the normal-state data (dashed line). Figure 2b compares σ_s/σ_n with $N_{\text{BCS}}(E)$ for $\Delta = 4.5$ meV (dashed line). The symmetric negative deviation of $\sim 1\%$ centred near 45 mV is characteristic of phonon effects¹. Inelastic tunnelling processes involving phonons produce increases in the tun-

nelling conductance near peaks in $F(\omega)$ (ref. 1). The lack of perfect symmetry in the deviation is attributed to the strong asymmetry of the normal-state (background) conductance (possibly owing to an asymmetric tunnel barrier), but features such as conductance dips are well correlated for both polarities. Equation (1) is strictly valid only for a constant background conductance and thus we always chose the polarity with the weakest voltage-dependent background for inversion, as is commonly done for tunnelling into conventional superconductors². Although the results might be different using the data with the sharper background, such a comparison has not been done.

We inverted the data to obtain $\alpha^2 F(\omega)$ using a modified McMillan-Rowell (MMR) procedure¹³. This procedure allows for a common problem in tunnelling: the presence of a thin proximity layer of reduced superconductivity on the surface, which diminishes the effect of phonons on σ_s . Studies¹³ of Nb_3Sn required MMR analysis to avoid anomalous values of λ and μ^* found with MR. Here we obtained convergence (after eight iterations) and a good fit of the data by assuming an energy-independent reduction of 10–25% in the measured phonon structure (in σ_s) owing to a proximity layer. Using the ordinary MR procedure resulted in a noticeable shift of the fitted tunnelling conductance compared to the data and an anomalous, negative value for μ^* . This shift is a signal of a proximity effect¹³ and the MMR procedure removes the shift completely, resulting in a much better fit of the data and correspondingly improved values for λ , μ^* and T_c .

The $\alpha^2 F(\omega)$ are shown in Fig. 3 for two BKBO (with $x = 0.375$) junctions, together with the generalized phonon density of states, $G(\omega)$ (for BKBO with $x = 0.4$), obtained by neutron scattering. The resolution of $G(\omega)$ is improved over previous data¹⁴. As $G(\omega)$ contains neutron-scattering cross-sections, it differs in its detailed shape from $F(\omega)$, but peaks and valleys occur at the same energies in both. Similarly, because α^2 may depend on ω , the shape of $\alpha^2 F(\omega)$ may be different from $F(\omega)$, but features such as peaks and valleys should occur at nearly the same energies. A good correlation of peaks in $G(\omega)$ and the two $\alpha^2 F(\omega)$ functions is seen for BKBO, and all show pronounced minima near 20 and 46 meV. The principal differences in the $\alpha^2 F(\omega)$ may be related to K concentration: the junction with $\Delta = 3.8$ meV has a stronger peak at 32 meV, whereas the junction with $\Delta = 4.5$ meV has a stronger peak at 10 meV. The total electron-phonon coupling strength is determined from an integral of $\alpha^2 F(\omega)$, and values of $\lambda = 1.2 \pm 0.2$ and 1.3 ± 0.3 are obtained for the BKBO junctions with $\Delta = 3.8$ and 4.5 meV, respectively. Uncertainties in λ are estimates based on the inversion analysis.

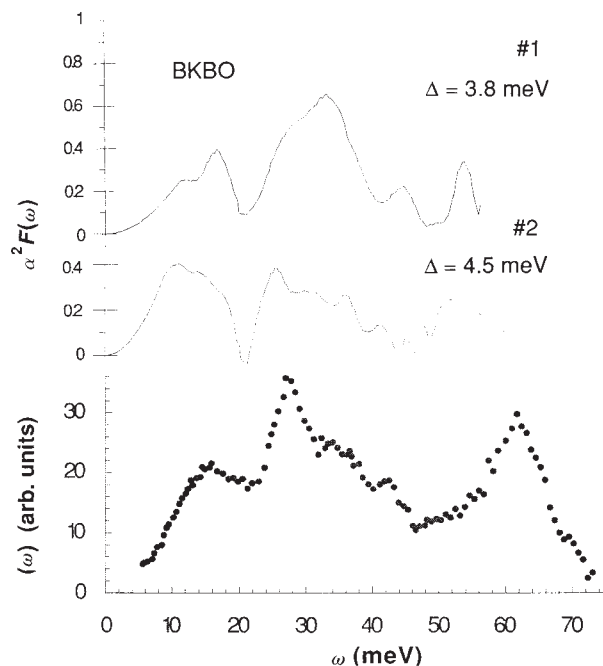


FIG. 3 The $\alpha^2 F(\omega)$ for two BKBO (with $x = 0.375$) junctions, and the generalized phonon density of states, $G(\omega)$ (for BKBO with $x = 0.4$), obtained by neutron scattering.

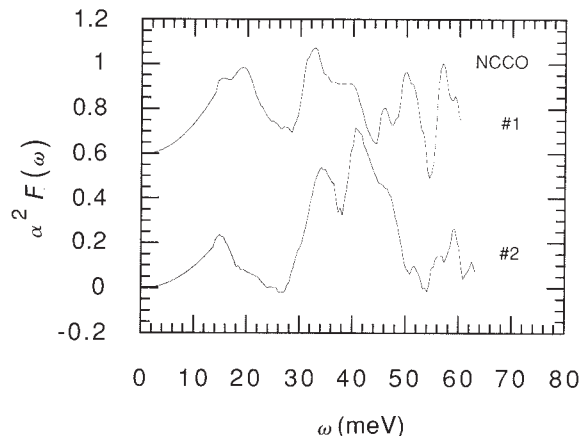


FIG. 4 The $\alpha^2 F(\omega)$ for two junctions on NCCO. The zero of junction 1 has been shifted by 0.6 for clarity.

For NCCO, we know of no measurements of $G(\omega)$; the $\alpha^2 F(\omega)$ for two junctions with $\Delta = 3.7$ meV are shown in Fig. 4. The correlation of peaks and valleys is good, and the two strong minima near 28 and 54 meV are similar to those in BKBO. Although $\alpha^2 F(\omega)$ drops below zero at 54 meV for junction 1, it is probably an artefact of the inversion owing to the poor signal-to-noise ratio above ~ 50 mV. The differences in the shapes may be due to anisotropy effects, which are expected to be larger in NCCO owing to the two-dimensional character of the Cu-O planes. Such anisotropy effects have been observed¹⁵ in the $\alpha^2 F(\omega)$ of conventional superconductors such as Nb.

A comparison of the tunnelling inversion with experimental data is given in Table 1. For the range of parameters leading to successful fits of the data, the calculated T_c changed by $\sim 10\%$, although the uncertainty in λ and μ^* was larger. The choice of cutoff of the inversion is near 65 meV, owing to a poor signal-to-noise ratio above that energy, which means that any potential contributions from higher-energy phonons or other mechanisms is left out. The T_c values calculated from $\alpha^2 F(\omega)$ are quite close to the measured ones for both materials, however, indicating that any such contributions may be unimportant. Also, the calculated λ and measured $2\Delta/kT_c$ are consistent with trends found in low- T_c superconductors. The uncertainty in λ and μ^* allows the possibility of another pairing mechanism, perhaps

as high as 25% of the phonon contribution. Nevertheless, the data suggest that phonons provide the predominant mechanism of superconducting pairing in NCCO and BKBO, within the framework of the Eliashberg theory, and they are thus similar to conventional superconductors. $\square$

Received 1 August; accepted 28 August 1990.

1. Wolf, E. L. *Principles of Electron Tunneling Spectroscopy* Chs 2–5 (Oxford University Press, New York, 1985).
2. Wolf, E. L., Zasadzinski, J., Osmun, J. W. & Arnold, G. B. *J. low Temp. Phys.* **40**, 19–50 (1980).
3. McMillan, W. L. & Rowell, J. M. in *Superconductivity* (ed. Parks, R. D.) Ch. 11 (Dekker, New York, 1969).
4. Zasadzinski, J. F. *et al. Physica C* **158**, 519–524 (1989).
5. Peng, J. L., Shelton, R. N. & Radousky, H. B. *Solid St. Commun.* **71**, 479–483 (1989).
6. Hinks, D. G., Mitchell, A. W., Zheng, Y., Richards, D. R. & Dabrowski, B. *Appl. Phys. Lett.* **54**, 1585–1587 (1989).
7. Huang, Q., Zasadzinski, J. F., Gray, K. E., Liu, J. Z. & Claus, H. *Phys. Rev.* **B40**, 9366–9369 (1989).
8. Huang, Q., Zasadzinski, J. F., Gray, K. E., Bukowski, E. D. & Ginsberg, D. M. *Physica C* **161**, 141–144 (1989).
9. Huang, Q., Zasadzinski, J. F. & Gray, K. E. *Phys. Rev.* (in the press).
10. Ekino, T. & Akimitsu, J. *Phys. Rev.* **B40**, 7364–7367 (1989).
11. Wang, S. H. *et al. Phys. Rev. Lett.* **64**, 1067–1070 (1990).
12. Schlesinger, Z. *et al. Phys. Rev.* **B40**, 6862–6866 (1989).
13. Wolf, E. L. *et al. Phys. Rev.* **B22**, 1214–1217 (1980).
14. Loong, C. K. *et al. Phys. Rev. Lett.* **62**, 2628–2631 (1989).
15. Wolf, E. L., Chen, T. P. & Burnell, D. M. *Phys. Rev.* **B31**, 6096–6098 (1985).

ACKNOWLEDGEMENTS. We thank Z. Li for her help in the sample preparation. This work was supported by the US Department of Energy and the NSF.

Effect on global warming of wind-dependent aerosol generation at the ocean surface

J. Latham & M. H. Smith

Department of Pure and Applied Physics,
University of Manchester Institute of Science and Technology,
Manchester M60 1QD, UK

CONSIDERABLE effort is being devoted^{1,2} to elucidating the influence of clouds on climate, and in particular to assessing the influence on global warming of modifications to cloud cover or microphysical characteristics. Such assessments are hindered by the present poor understanding of the many feedback processes involved. Here we demonstrate the existence of a negative feedback process in which an increase in wind speed (a possible result of atmospheric warming) produces, over oceanic regions, increased emissions of sea-salt aerosol particles, which subsequently act as cloud condensation nuclei. Activation of these nuclei could then significantly increase the number concentration of cloud droplets in marine stratus, thereby enhancing the albedo of these clouds for incoming short-wave radiation and so producing a cooling effect. Our calculations indicate that an increase in wind speeds of about $5\text{--}10\text{ m s}^{-1}$ would produce an increase in cloud albedo² sufficient to compensate for predicted levels of global warming.

It is well established^{3,4} that aerosols are generated at the ocean surface by wind-related processes such as bubble-film bursting, whitecapping and spin-drift. The generation of such aerosol particles increases rapidly with increasing wind speed U (refs 5–7). The primary constituent of this aerosol is sodium chloride, which is commonly found in chemical analyses of marine cloud water. Current opinion⁸, however, is that although a small fraction of the cloud condensation nuclei (CCN) on which maritime cloud droplets are formed probably consists principally of NaCl, most natural CCN over the oceans consist of ammonium sulphate particles formed from dimethylsulphide produced at the ocean surface by planktonic algae.

Thus, although the generation of sea-salt aerosol particles at the ocean surface by bubble-bursting and the other processes (1) is an important source of giant nuclei, which might facilitate precipitation development in cumulus clouds⁹; (2) has important

influences (of possible climatological significance) on the optical characteristics of the marine atmosphere; and (3) can play an important part in moisture exchange and energy transport between the ocean and the atmosphere, it seems to make only a secondary contribution to the formation of droplets in marine stratus cloud.

It is possible, in principle, that if increased wind speeds accompany global warming, the droplet number concentration N will increase accordingly, to a sufficient extent to produce a significant increase in cloud albedo. It has been calculated² that the warming associated with a doubling of the atmospheric concentration of CO_2 can be counteracted by a decrease of 10–15% in effective radius of the droplets, which corresponds to an increase of $\sim 30\text{--}45\%$ in the number concentration.

We made aerosol measurements in the North Atlantic, in all seasons, from a promontory on the northwest tip (Ardvachar Point) of the island of South Uist in the Outer Hebrides. The range of wind speeds covered was $\sim 0\text{--}30\text{ m s}^{-1}$, and the concentration of aerosol n and volumetric loading V increased with wind speed for all aerosol sizes d examined, which ranged from about 0.2 to $100\text{ }\mu\text{m}$, the rate of increase being greater as d increased.

Figure 1 illustrates the observed increase of V with U for particles in the range $16\text{--}32\text{ }\mu\text{m}$. The curve through the experimental points is described by the equation

$$\ln V \approx 0.2U + 4 \quad (1)$$

where V is in $\mu\text{m}^3\text{ cm}^{-3}$ and U is in m s^{-1} . A similar expression describes the variation of aerosol number concentration with wind speed.

Although aerosol in this size range would be very effective CCN, their concentrations are far too low, particularly at cloud altitudes, to exercise a significant influence on N . Probably the most important size range of sea-salt aerosol particles is $\sim 0.5\text{--}1\text{ }\mu\text{m}$. Such particles are large enough to be more effective CCN than those originating from dimethylsulphide, but small enough to be present in appreciable concentrations and distributed reasonably uniformly throughout the well mixed boundary layer over the oceans. For aerosols in this size range⁷

$$\log n \approx 0.1U - 0.3 \quad (2)$$

where n is in cm^{-3} (ref. 7).

If we assume that the concentration of CCN given by equation (2) corresponds to the observed mean ocean wind speed in the

FIG. 1 Variation of the aerosol volumetric loading V with wind speed U . The open triangles denote data from ref. 6, the crosses from ref. 11 and the solid squares from ref. 7.

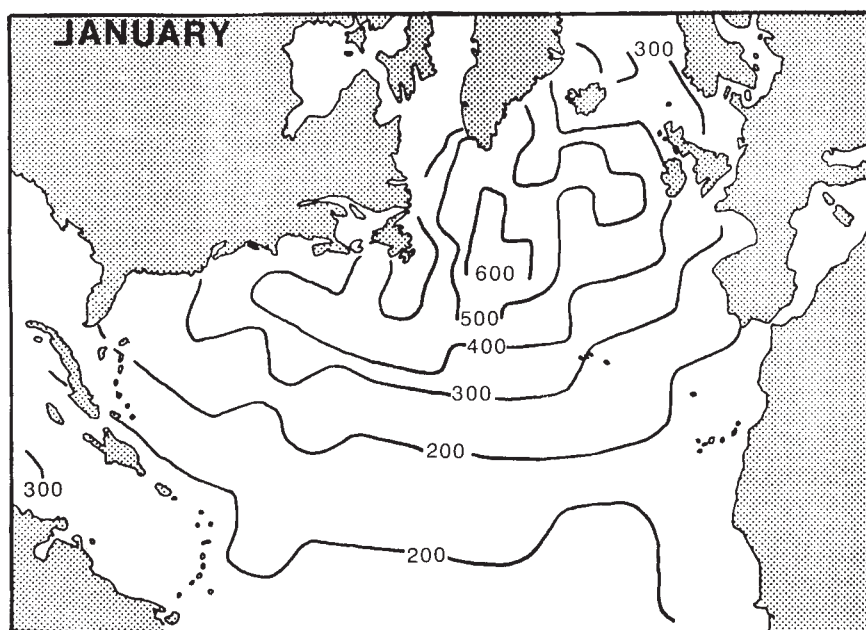
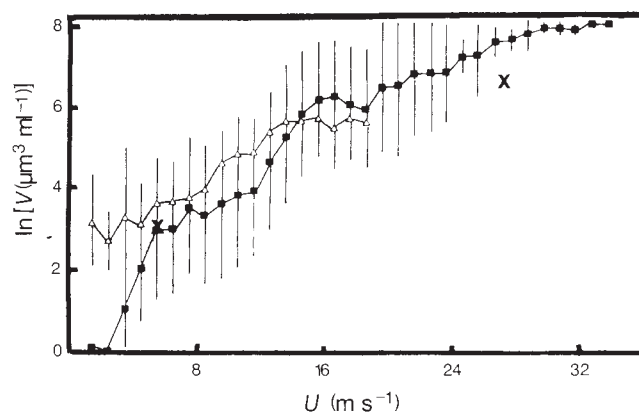


FIG. 2 Distribution of calculated mean surface volumetric aerosol particle loadings ($\mu\text{g m}^{-3}$) for the North Atlantic in January, based on monthly averaged values of 20 years of North Atlantic windspeeds, and using the relationship¹² $\ln V = 0.17U + 4$.

North Atlantic of $\sim 10 \text{ m s}^{-1}$, the associated concentration of cloud droplets is $\sim 5 \text{ cm}^{-3}$, a small fraction of the overall droplet concentration, as earlier research indicated¹. But if, as a consequence of global warming or any other process, the average value of U increases to 20 m s^{-1} , N will increase to 50 cm^{-3} , and if U increases to 30 m s^{-1} , N will increase to $\sim 500 \text{ cm}^{-3}$. A typical value of total droplet concentration for marine stratus is 50 cm^{-3} , and thus we predict that increases of $\sim 10 \text{ m s}^{-1}$ in average wind speeds may sufficiently enhance droplet populations in marine stratus to produce a cooling comparable² with the estimated global warming. If other estimates¹⁰ of characteristic concentrations of sea-salt CCN are correct, significant cooling could result from an increase in U of $\sim 5 \text{ m s}^{-1}$.

The foregoing arguments are highly simplistic. Additional sea-salt CCN produced at the ocean surface may, because their activation thresholds are lower, inhibit the activation of CCN produced from dimethylsulphide, thereby causing an overall reduction in N . In this case the albedo for short-wave radiation will decrease. Also, an increase in wind speed could produce

an increase in the production rate of dimethylsulphide, and thus in the concentrations of CCN. Because the sea-salt aerosols are more readily activated than those produced from dimethylsulphide, this factor is unlikely to influence significantly the foregoing arguments. Finally, the distribution of wind speed over the oceans is distinctly non-uniform. Figure 2 shows predicted variations of volumetric aerosol loading over the North Atlantic during January, for the size range $0.5 < d < 32 \mu\text{m}$. The nonlinearity of the n - U relationship must be taken into account, and it seems likely that some localized regions experiencing especially high wind speeds may act as sources of CCN for much more extensive regions.

This sea-salt aerosol generation process needs to be considered a possibly important feedback process associated with global warming; and should therefore be incorporated into climate models. Our ideas could be assessed, in principle, by examining satellite data on cloud albedos, with associated surface wind measurements. □

Received 27 June; accepted 13 August 1990.

- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655-661 (1987).
- Slingo, A. *Nature* **343**, 49-51 (1990).
- Blanchard, D. C. *J. Rech. Atmos.* **4**, 1-6 (1969).
- Monahan, E. C., Fairall, C. W., Davidson, K. L. & Boyle, P. J. *Q. J. R. Met. Soc.* **109** (1983).
- Wells, W. C., Gal, G. & Munn, M. W. *Appl. Opt.* **16**, 654-659 (1977).
- Exton, H. J. *et al. Q. J. R. Met. Soc.* **111**, 817-837 (1985).
- Smith, M. H., Consterdine, I. E. & Park, P. M. *Q. J. R. Met. Soc.* **115**, 383-396 (1989).

- Bigg, E. K., Gras, J. L. & Evans, C. J. *atmos. Chem.* **1**, 203-214 (1984).

- Illingworth, A. J. *Nature* **336**, 754-756 (1988).

- Blanchard, D. C. & Capriano, R. *Nature* **330**, 526-528 (1987).

- Woodcock, A. H. *J. geophys. Res.* **77**, 5316-21 (1950).

- Exton, H. J., Latham, J., Park, P. M., Smith, M. H. & Allan, R. R. in *Oceanic Whitecaps* (eds Monahan, E. C. & MacNiocaill, H.) 175-193 (Reidel, Dordrecht, 1983).

ACKNOWLEDGEMENTS. This work was supported by the Ministry of Defence and the Hadley Centre for Climate Prediction and Research.

Climatically induced rapid acidification of a softwater seepage lake

K. E. Webster*, A. D. Newell†, L. A. Baker‡
& P. L. Brezonik§

* Wisconsin Department of Natural Resources, 3911 Fish Hatchery Road, Madison, Wisconsin 53711, USA

† NSI Technology Services and ‡ Water Resources Research Center (University of Minnesota), EPA Environmental Research Laboratory, 200 SW 35th Street, Corvallis, Oregon 97333, USA

§ Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA

To establish a relationship between levels of atmospheric pollutants such as SO_2 and the acid-base chemistry of lakes is an important but elusive goal of studies on acid precipitation. But the direct effect of acid deposition on the acid-neutralizing capacity (ANC) of lake waters is difficult to isolate, because rates of change of the latter are small¹ and ANC is also influenced by other factors^{2,3}. We report here the observation of changes in the ANC of a seepage lake in Michigan driven by the effects of drought on the local hydrology. While levels of acidic wet deposition at the lake remained at a constant, moderate level over a five-year period, a rapid, major loss of ANC occurred because of changes in the input of ANC-rich groundwater. Thus even modest climate fluctuations can have substantial effects on lake chemistry, and must be taken into account when interpreting temporal trends. Sustained drought conditions might similarly alter the acid-base chemistry of other softwater systems.

The ANC of softwater lakes is determined by four main factors: the geological setting, the hydrological paths of inputs, internal alkalinity generation, and the atmospheric loading rates of mineral acids. Seepage lakes are defined hydrologically by the absence of permanent inlets and outlets of surface water, and their sensitivity toward acidification is closely related to the relative amount of ANC-rich groundwater they receive³. For many of the low-ANC seepage lakes that receive $\leq 10\%$ of their water from groundwater, internal alkalinity-generating processes such as sulphate reduction become an important source of ANC⁴. Seepage lakes occur worldwide, commonly in glacial outwash deposits, coastal dunes and karst topography. In geologically sensitive (that is, slowly weatherable) regions of

the United States sampled during the National Surface Water Survey, $>25\%$ of the lakes were seepage lakes. More important from an impact perspective, 56% of the acidic lakes ($\text{ANC} < 0 \mu\text{eq l}^{-1}$) and 41% of the lakes with $\text{ANC} < 50 \mu\text{eq l}^{-1}$ were seepage systems⁵.

Nevins Lake ($46^\circ 31' 00'' \text{ N}$, $86^\circ 14' 34'' \text{ W}$) is a small (110 ha), continually mixed seepage lake (maximum and mean depths of 4.9 and 1.8 m) located in the eastern Upper Peninsula of Michigan. The lake lies on a highland glacial outwash plain of sand and gravel⁶ within 10 km of Lake Superior and is surrounded by white pine and hardwood forest. The watershed is relatively remote, but 30 seasonal cottages and one year-round residence are present on the northern shore. The region receives $\sim 95 \text{ cm yr}^{-1}$ of precipitation, and the influence of Lake Superior is reflected in the wide range of snowpack measurements from two nearby stations (530–627 cm).

As part of a continuing study designed to detect trends related to acid deposition⁷, water chemistry variables were analysed between autumn 1983 and autumn 1988 on samples collected at 1.0 m depth each spring (after the ice thawed), mid-summer and autumn. To evaluate time trends, we applied the seasonal Kendall τ -test (with correction for serial dependence)⁸ to the SO_4^{2-} , Ca^{2+} , Mg^{2+} , ANC, pH, conductance and dissolved organic carbon (DOC) data. For variables showing significant trends, the median rate of change was calculated using the seasonal Kendall slope estimator⁹.

During the study period, there was a striking change in the acid-base chemistry of the lake. The ANC decreased by a factor of 8 from a peak concentration of $178 \mu\text{eq l}^{-1}$ in autumn 1983 to a low of $22 \mu\text{eq l}^{-1}$ in autumn 1988 (Fig. 1), a highly significant

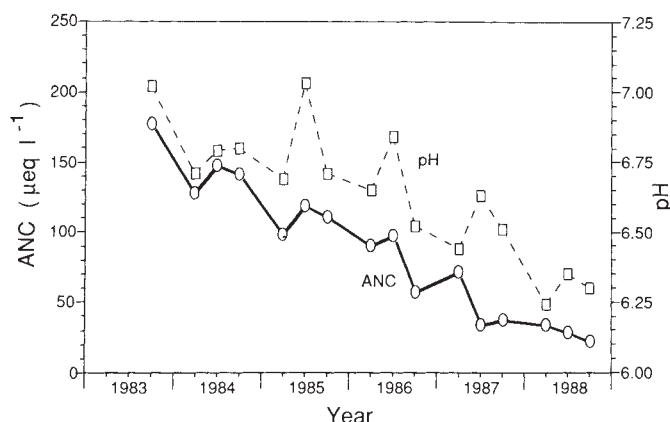


FIG. 1 ANC and pH measured in samples collected at 1.0 m depth from Nevins Lake between autumn 1983 and autumn 1988. Trend slopes were $-30 \mu\text{eq l}^{-1} \text{ yr}^{-1}$ for ANC (95% confidence interval -17 to $-40 \mu\text{eq l}^{-1} \text{ yr}^{-1}$; $p=0.007$) and -0.14 units yr^{-1} for pH (95% confidence interval -0.04 to -0.21 ; $p=0.011$). ANC was determined by Gran titration to pH 3.5 and pH samples were not aerated.

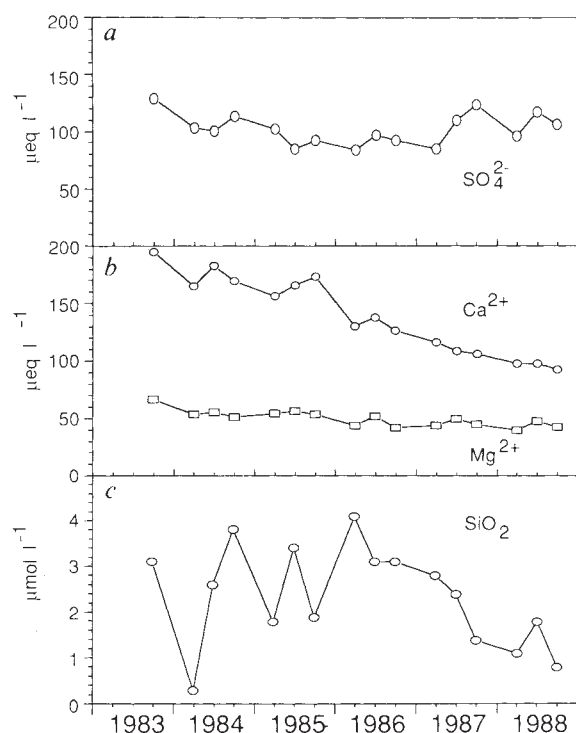


FIG. 2 Concentrations of a, SO_4^{2-} , b, Ca^{2+} and Mg^{2+} and c, SiO_2 in samples collected at 1.0 m depth from Nevins Lake. Trend slopes were $-20 \mu\text{eq l}^{-1} \text{ yr}^{-1}$ for Ca^{2+} (95% confidence interval -12 to $-28 \mu\text{eq l}^{-1} \text{ yr}^{-1}$; $p=0.010$), $-3 \mu\text{eq l}^{-1} \text{ yr}^{-1}$ for Mg^{2+} (95% confidence interval -1 to $-5 \mu\text{eq l}^{-1} \text{ yr}^{-1}$; $p=0.028$), and $-23 \mu\text{eq l}^{-1} \text{ yr}^{-1}$ for $\Sigma \text{Ca}^{2+} + \text{Mg}^{2+}$ (95% confidence interval -16 to $-32 \mu\text{eq l}^{-1} \text{ yr}^{-1}$; $p=0.011$). No significant trend was detected for SO_4^{2-} ($p=0.641$). Analytical methods: SO_4^{2-} , ion chromatography; base cations (dissolved), atomic absorption spectrophotometry; SiO_2 , oxalic acid/molybdate method.

change ($-30 \mu\text{eq l}^{-1} \text{yr}^{-1}$). This was mirrored by a significant decrease in the pH from 7.03 to 6.24 (Fig. 1). Parallel decreases in Ca^{2+} and Mg^{2+} (19 and $3 \mu\text{eq l}^{-1} \text{yr}^{-1}$, respectively) accounted for most of the loss of ANC, with the balance attributable either to variance about these slope estimates or to small undetected trends in other variables. Of the remaining variables, a significant trend was detected only for conductance ($-2.3 \mu\text{S cm}^{-1} \text{yr}^{-1}$, $p = 0.007$, where P is the probability that $\tau = 0$); DOC, Na^+ , K^+ and Cl^- remained relatively constant at low concentrations.

Several lines of evidence support our hypothesis that the acidification (loss of ANC) was driven by fluctuations in climate and hydrology, and not by increases in atmospheric pollutant loading, shifts in internal alkalinity-generating processes, or watershed disturbance. Nevins Lake receives moderate loadings of mineral acids; based on interpolation of data from the National Atmospheric Deposition Program, wet SO_4^{2-} deposition is $15.1 \text{ kg ha}^{-1} \text{yr}^{-1}$ and the volume-weighted pH of the precipitation is 4.6 (T. Olsen, personal communication). Wet-deposition rates of base cations and strong-acid anions remained stable between 1979–1987 (seasonal Kendall τ -test) at the closest NADP stations (Trout Lake, Wisconsin and Douglas Lake, Michigan)¹⁰. Further, we observed no significant trend in lake SO_4^{2-} during the study period (Fig. 2a). Nitrate was detectable only in spring samples and then concentrations were low ($<10 \mu\text{eq l}^{-1}$). The magnitude of ANC loss, combined with stable SO_4^{2-} concentration in the lake, discount a change in internal alkalinity generation by sulphate reduction as the primary explanation for the patterns we observed. Low concentrations of soluble reactive phosphorus ($<3 \mu\text{g l}^{-1}$) and the lack of change in land use during the study period suggest that shoreline dwellings were not a factor inducing the chemical changes.

The monotonic decreases in Ca^{2+} , Mg^{2+} and SiO_2 (Fig. 2b, c) strongly suggest that the rapid acidification of Nevins Lake was linked to reduced groundwater discharge. Superimposed on the progressive loss of ionic strength, we observed summer maxima in ANC and Ca^{2+} (Figs. 1, 2b) which may reflect pulsed inputs of groundwater following the melting of the snow. The absence of this seasonal pattern after 1986, concurrent with a decrease in SiO_2 (Fig. 2c), probably marks the end of the flow of groundwater into the lake. Low-ANC seepage lakes receive a large portion of their ANC, Ca^{2+} , Mg^{2+} and SiO_2 from groundwater, although their hydrological inputs are dominated by precipitation rather than groundwater^{11–14}. Elimination or reduction of groundwater discharge combined with continued inputs of

mineral acids from precipitation, gradually depletes the ANC, primarily through dilution^{12,15}. Ion-enrichment calculations¹¹ predict that prolonged absence of groundwater inputs to Nevins Lake would eventually result in acidic conditions ($\text{ANC} \leq 0$).

Climate data further support our interpretation of the chemical trends. Five-year running averages of annual precipitation at Grand Marais (22 km to the northeast of the study site) were far below normal during the entire study period (Fig. 3). Hydrological records are unavailable, but the lake level in 1988 was as much as 0.5 m below that of previous years (local residents' and our observations), corresponding to $\sim 10\%$ loss of volume. In addition, the water table at three US Geological Survey monitoring wells 13 to 95 km from Nevins Lake had decreased significantly (seasonal Kendall τ -test) since 1975.

The speed and magnitude of a lake's chemical response to drought seems to be related to the initial percentage of groundwater input, which determines ANC, and water residence time. During the same period, drought-related decreases in the lake level and in the water table, followed by cessation of groundwater inflow, were observed at Vandercook Lake and Little Rock Lake, two low-ANC ($20\text{--}25 \mu\text{eq l}^{-1}$) seepage lakes in north central Wisconsin (D. A. Wentz and W. J. Rose, personal communication). During this time, however, a drought-related decrease in ANC of $10 \mu\text{eq l}^{-1}$ occurred only in the untreated basin of Little Rock Lake (C. Watras, personal communication). This decrease in ANC was much less than that observed at Nevins Lake ($150 \mu\text{eq l}^{-1}$) owing to Little Rock Lake's lower initial ANC ($25 \mu\text{eq l}^{-1}$), longer residence time (~ 3 yr) and smaller groundwater contribution. The relatively short inflow-based water residence time (~ 1.6 yr) of Nevins Lake, the result of a shallow mean depth combined with large annual inputs of precipitation, explains its rapid response. Lakes with longer water residence times, such as Vandercook Lake (~ 5 yr)¹⁶, would be expected to exhibit lags in their chemical response to hydrological fluctuations.

The data from Nevins Lake demonstrate that changes in hydrological flowpaths caused by climate fluctuations—in this case a shift in the relative inputs of groundwater and direct precipitation—can cause ANC changes exceeding those expected in response to acidic deposition^{1,17}. It is likely that the chemical changes in Nevins Lake represent an extreme in terms of the magnitude and the speed of the response. Nevertheless, the process we describe is generally applicable to systems that rely on small inputs of groundwater to sustain their ANC, especially low-ANC seepage lakes and peatlands¹⁸. Uncertainties about the effects of drought on overland and stream flow (hydrological inputs absent from seepage lakes) complicate but do not preclude extrapolation to low-ANC drainage lakes. Our results suggest that reduced groundwater inputs related to sustained drought or to global warming increase the risk of acidification to many softwater systems. □

Received 30 April; accepted 10 August 1990.

1. Schindler, D. W. *Science* **239**, 149–157 (1988).
2. Eilers, J. M., Glass, G. E., Pollack, A. K. & Sorenson, J. A. *Can. J. Fish. aquat. Sci.* **46**, 1929–1944 (1989).
3. Eilers, J. M., Glass, G. E., Webster, K. E. & Rogalla, J. A. *Can. J. Fish. aquat. Sci.* **40**, 1896–1904 (1983).
4. Baker, L. A. & Brezonik, P. L. *Wat. Resour. Res.* **24**, 65–74 (1988).
5. Baker, L. A., Kaufmann, P. R., Herlihy, A. T. & Eilers, J. M. *Current Status of Surface Water Acid-Base Chemistry, Rep. SOS/T 9* (National Acid Precipitation Assessment Program, Washington, DC, 1990).
6. Vanlier, K. E. *Reconnaissance of the Ground-water Resources of Alger County, Michigan, Water Investigation 1* (Michigan Department of Conservation, Geological Survey Division, Lansing, 1963).
7. Newell, A. D., Powers, C. F. & Christie, S. J. *Rep. EPA/600/4-87/014*. (US Environmental Protection Agency, Washington, DC, 1987).
8. Hirsch, R. M. & Slack, J. R. *Wat. Resour. Res.* **20**, 727–732 (1984).
9. Hirsch, R. M., Slack, J. R. & Smith, R. A. *Wat. Resour. Res.* **18**, 107–121 (1982).
10. *NADP/NTN Annual Data Summaries 1980–1987* (Colorado State University, Fort Collins, 1981–1988).
11. Baker, L. A., Eilers, J. M., Cook, R. B., Kaufmann, P. K. & Herlihy, A. H. in *Acidic Deposition and Aquatic Ecosystems* (ed. Charles, D. F.) (Springer-Verlag, New York, in the press).
12. Anderson, M. P. & Bowser, C. J. *Wat. Resour. Res.* **22**, 1101–1108 (1986).
13. Wentz, D. A. *et al.* in *Acid Rain* (eds Perry, R., Harrison, R. M., Bell, J. N. B. & Lester, J. N.) 309–316 (Selper, London, 1987).
14. Hurley, J. P., Armstrong, D. E., Kenoyer, G. J. & Bowser, C. J. *Science* **227**, 1576–1578 (1985).

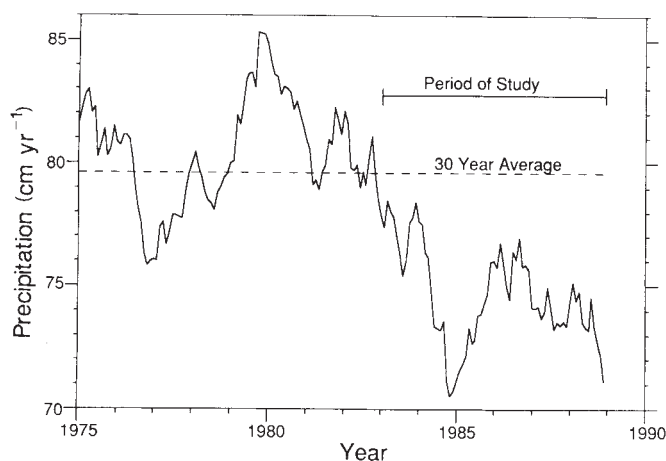


FIG. 3 Five-year running means of annual precipitation at Grand Marais, calculated monthly. The dashed line indicates the 30-yr average annual precipitation at the site. NOAA climatological data 1970–88 (National Climatic Center, Asheville, North Carolina).

15. Garrison, P. J. *et al.* *Lake Reservoir Management* **3**, 356–364 (1987).
16. Wentz, D. A. & Rose, W. J. *Arch. envir. contam. Tox.* **18**, 147–155 (1989).
17. Dillon, P. J., Reid, R. A. & de Grosbois, E. *Nature* **329**, 45–48 (1987).
18. Gorham, W., Bayley, S. E. & Schindler, D. W. *Can. J. Fish. aquat. Sci.* **41**, 1256–1268 (1984).

ACKNOWLEDGEMENTS. We thank D. G. Soltis, M. D. Johnson, B. J. Holdhusen, J. V. Adams and R. C. Hjort for their contributions, and J. L. Stoddard, T. M. Frost, T. C. Young and J. M. Eilers for reviews. We also thank the US Geological Survey for data on groundwater levels and F. V. Nurnberger for snowpack data. The work was supported by the US Environmental Protection Agency.

Emission of electromagnetic radiation preceding the Ito seismic swarm of 1989

Yukio Fujinawa* & Kozo Takahashi†

* National Research Institute for Earth Science and Disaster Prevention, Tennodai 3-1, Tsukuba-shi, Ibaraki-ken, 305, Japan

† Communications Research Laboratory, Nukukita 4-2-1, Koganei-shi, Tokyo, 184, Japan

THERE have been several reports^{1–8} of electromagnetic radiation being detected in air and/or underground before large earthquakes or volcanic eruptions. Unfortunately, such measurements are often plagued by contamination from other sources of noise. We have developed an electrode system that is effective at filtering out urban and atmospheric background noise. Here we report the detection of anomalous electromagnetic signals several hours before the occurrence of large earthquakes near Ito city, Japan, about 150 km from the location of the electrodes. We have also detected anomalous signals about one day before an undersea volcanic eruption that constituted part of the same seismic swarm near Ito. Although the mechanism responsible for such signals is still unclear, these results indicate that their monitoring could be valuable for the prediction of seismic activity.

Conventional methods of measuring underground electromagnetic radiation (EMR) associated with seismicity are not effective in a highly industrialized area with high levels of urban noise. Measuring EMR using a steel pipe in a deep borehole as an electrode was therefore proposed, especially to measure the vertical component of the underground electric field⁹. This configuration is very effective for reducing signals caused by lightning discharges or other atmospheric effects⁹. Such signals

propagate between the ionosphere and the ground in a 'waveguide mode', which means that the associated electric fields under the ground surface are predominantly horizontal. The electrode system was installed in late March 1989 at Tsukuba, ~60 km northeast of Tokyo. One electrode is a steel pipe in a borehole 603 m deep and the other is a circle of grounded wire, 40 m across, which surrounds the hole at a depth of 1 m. Although we have no means of knowing the exact extent of earthing of the casing pipe at an arbitrary depth, we suspect that the finite value of earth resistivity would cause partial grounding. At Tsukuba, the surface of the pre-Tertiary basement¹⁰ is at a depth of 400 m. Below the ground surface the resistivity is several tens of ohm metres, but lower down, it is an order of magnitude more conductive owing to the high salinity of fossil water¹¹. The lowest layer below the basement surface has a high resistivity of several hundreds of ohm metres. The vertical inhomogeneity in the resistivity enables the electrode system to be sensitive to vertical gradients of the electric field.

We have observed EMR in three frequency bands—the d.c. range (~12 Hz), the ultra-low-frequency (ULF) range (0.01–~12 Hz), and the extremely-low-frequency (ELF) and very low-frequency (VLF) ranges (1–~9 kHz)—since last March. We detected no anomalous record in the ULF ranges with our recording method, however, and the following discussion is limited to the ELF/VLF bands. Amplitude envelopes of the ELF/VLF band have been recorded on strip charts. Criteria for determining abnormal radiation were tentatively selected by examining the normal patterns before and after events.

The most anomalous precursory EMR was observed during the seismic swarm and the undersea volcanic eruption that took place in 1989 off the coast of Ito (Fig. 1)¹². (Seismic swarms have been occurring almost annually in the western area of Sagami Bay since 1978¹³.) Last summer's swarm activity started at the end of June and on 5 July 1989 a moderate earthquake of magnitude $M = 4.9$ occurred near Ito, at 02:28 local time. Figure 2a shows the EMR amplitude from ~16:00 on 4 July to ~05:00 on 5 July. At ~20:00 the radiation pattern suddenly changed; the first burst in radiation level was maintained for half an hour and in the second burst the level was maintained for about an hour. These bursts occurred ~6 h and 4 h before the earthquake, respectively. Typical radiation patterns of normal conditions are shown in Fig. 2b. Normal patterns consist almost entirely of short-lived pulses, and are thought to be the result of atmospheric activity.

The anomalous radiation is characterized by sporadic bursts lasting half an hour, together with numerous large pulses. By inspecting the data with an event-recorder we found that the durations of the pulses were >50 ms, much longer than the ~5-ms duration of pulses caused by atmospheric activity. The pulses continued until the earthquake occurred. There were no appreciable abnormal fluctuations in the magnetic field data from Ohshima, which is in the vicinity of the seismic swarm at the time of the anomalous radiation.

The largest shallow earthquake ($M = 5.5$; focal depth, 17 km) during the swarm occurred at 11:09 on 9 July. The EMR recorded before this earthquake is shown in Fig. 3. Prominent, sporadic emission bursts, about half an hour long, occurred intermittently for ~10 h before the earthquake. Because 108 earthquakes were recorded on this day (Japan Meteorological Agency), it is not clear whether or not the bursts were related to a particular earthquake. The regularity of the bursts, however, suggests that the abnormal radiation was related to the largest earthquake.

The seismic swarm peaked between 4 and 10 July with prominent crustal motions, such as observed by the Global Positioning System (GPS) technique¹⁴. On 11 July a volcanic tremor started without any appreciable crustal motion, however, causing concern among the disaster prevention communities in Japan. Our EMR data showed a distinctive abnormal radiation pattern from ~13:00 to ~17:00 on the day before the eruption on 13 July (Fig. 4). The radiation pattern is obviously different from that

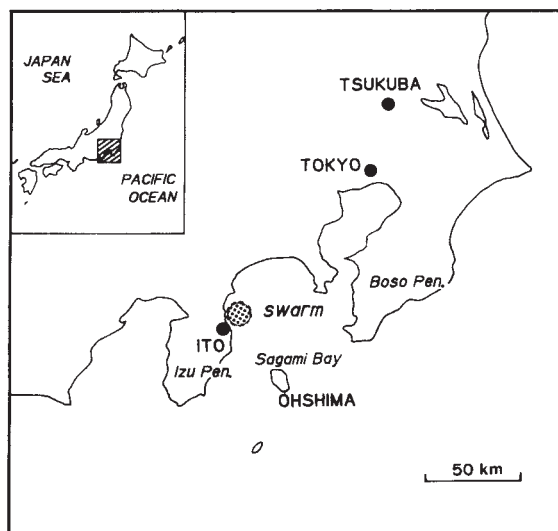


FIG. 1 The electromagnetic radiation recording system is installed at Tsukuba. The 1989 seismic swarm occurred off the coast of Ito on the Izu peninsula.

of a normal state and is also different from the radiation before the large earthquake on 9 July. The data is reminiscent of the abnormal EMR at 82 kHz observed a day before the 1986 eruption of Mt Mihara⁴. The anomalous signals continued for ~4 h, together with pulses similar to those before an earthquake, which were used to predict that eruption.

Laboratory experiments on rocks^{1,15} indicate that EMR is

caused by microscopic fractures. Other experiments¹⁶⁻¹⁸ on quartz-bearing and on non-quartz-bearing rock samples suggest that EMR is caused by positive charge accumulation and negative charge expulsion at fresh surfaces of tensile microfractures, and that the dominant mechanism for this electron expulsion owing to the opening of individual microfractures¹⁹ rather than piezoelectric effects¹. If we apply this hypothesis to explain the

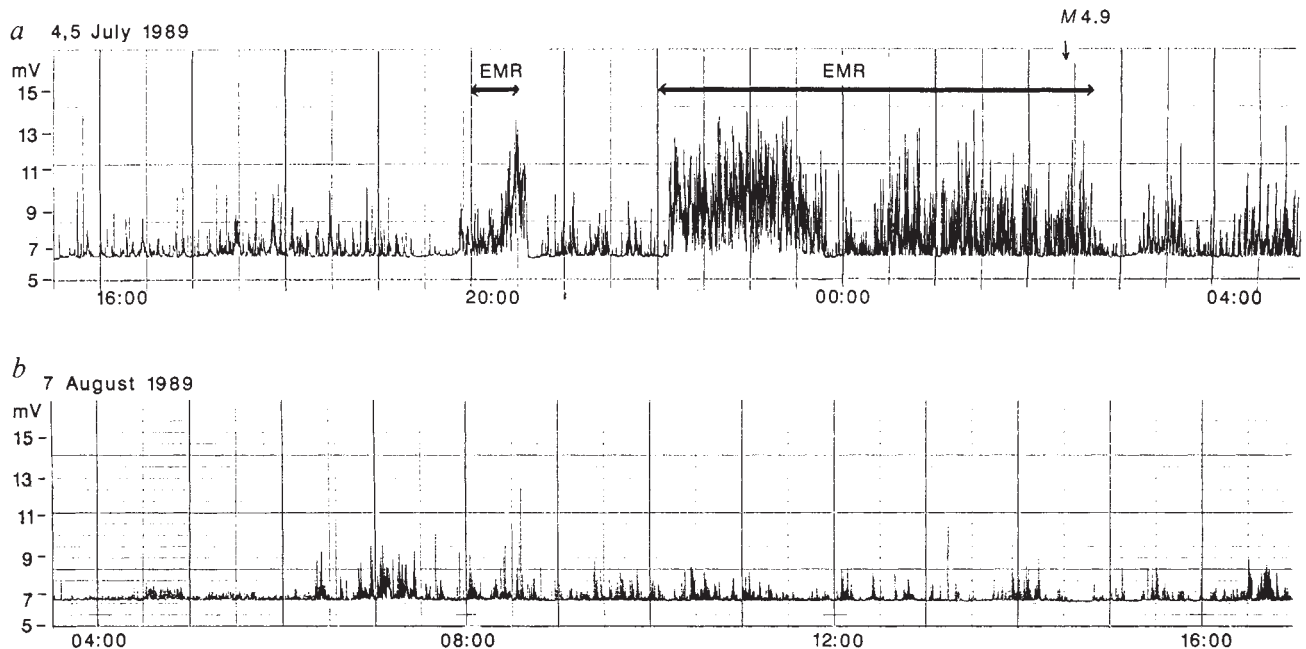


FIG. 2 *a*, electromagnetic radiation pattern before and after the earthquake ($M=4.9$) of 5 July 1989. The anomaly is characterized by radiation of relatively constant amplitude together with pulses. *b*, An amplitude record

of normal electromagnetic radiation, which is characterized by pulses owing to atmospheric effects.

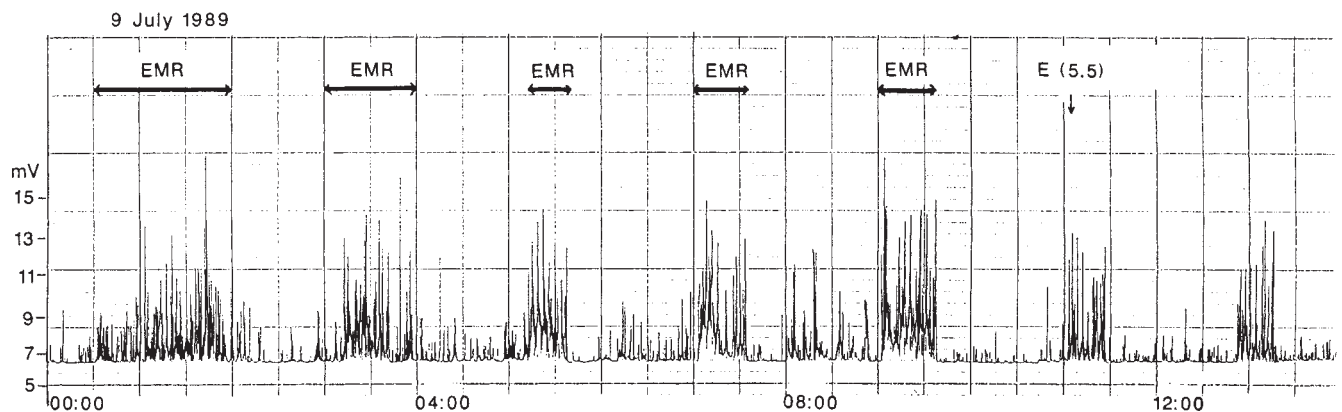


FIG. 3 Pattern of electromagnetic radiation before the earthquake ($M=5.5$) of 9 July 1989.

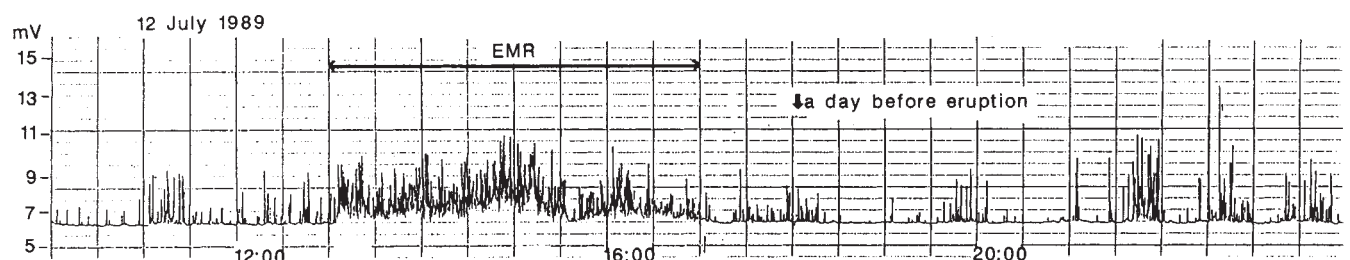


FIG. 4 Pattern of electromagnetic radiation before the undersea volcanic eruption of 13 July 1989.

intermittent sporadic EMR preceding the earthquake, propagation of microfracturing is expected to proceed, but to be intermittently obstructed by stronger parts of the crust²⁰. The long continuous EMR obtained before the volcanic eruption suggests that tensile cracks were produced continuously by dyke intrusion¹⁸.

Before crustal rupture, microscopic processes occur in most cases, usually being observed as acoustic emissions or electromagnetic emissions, and are more or less governed by anelastic or nonlinear properties²¹. These processes will be examined through EMR to study the source mechanisms of seismic ruptures and volcanic eruptions.

Several earthquakes with magnitudes >5.0 have occurred in central Japan since our observations began. Tentative statistics about the relation between EMR and these earthquakes revealed that a prominent EMR preceded these events by several hours to several days, when the earthquakes occurred in shallow depths of ~0–60 km inland or in the sea bed within ~300 km from the observation site⁵. Using EMR data, we hope to develop a practical and useful way of predicting shallow earthquakes several days in advance. An instrument similar to the one in Tsukuba has been installed on Ohshima (Fig. 1), about 160 km southwest of Tsukuba, and several others are planned for the central

part of Japan. These will increase the reliability of the observations, and make it possible to locate the source region from the phase difference of EMR signal at different sites⁹. □

Received 5 March; accepted 9 August 1990.

1. Warwick, J. W. *et al.* *J. geophys. Res.* **87**, 2851–2859 (1982).
2. Gokhberg, M. B. *et al.* *J. geophys. Res.* **87**, 7824–7828 (1982).
3. Varotsos, P. & Alexopoulos, K. *Tectonophysics* **110**, 73–98 (1984).
4. Yoshino, T. & Tomizawa, I. *Phys. Earth planet. Inter.* **57**, 32–39 (1989).
5. Oike, K. & Ogawa, T. *J. Geomagn. Geoelec., Kyoto* **38**, 1031–1040 (1986).
6. Tate, J. & Daily, W. *Phys. Earth planet. Inter.* **57**, 1–10 (1989).
7. Nikiforova, N. N. *et al.* *Phys. Earth planet. Inter.* **57**, 68–75 (1989).
8. Frazer-Smith, A. C. *et al.* *Abstr. Am. geophys. Un. Meeting December 1989*.
9. Takahashi, H. & Takahashi, K. *Phys. Earth planet. Inter.* **57**, 40–44 (1989).
10. Ikeda, R. *Rep. Coord. Comm. Earthquake Prediction* **32**, 157–158 (1984).
11. Suzuki, H. *et al.* *Res. Notes Natn. Res. Centre Disaster Prev.* No. 48 (1983).
12. *Nature* **340**, 175 (1989).
13. Ishida, M. *Proc. Earthquake Prediction Res. Symp.* 51–60 (1987) (in Japanese).
14. Shimada, S. *et al.* *Nature* **343**, 631–632 (1990).
15. Nitsan, U. *Geophys. Res. Lett.* **4**, 333–336 (1977).
16. Cress, G. O. *et al.* *Geophys. Res. Lett.* **14**, 331–334 (1987).
17. Yamada, I. *et al.* *Phys. Earth planet. Inter.* **57**, 157–168 (1989).
18. Ogawa, T. *et al.* *J. geophys. Res.* **90**, 9245–9249 (1985).
19. Kramer, J. *Acta phys. austriaca* **10**, 327 (1957).
20. Aki, K. & Richard, P. G. *Quantitative Seismology* (Freeman, San Francisco, 1980).
21. Tse, S. T. & Rice, J. R. *J. geophys. Res.* **91**, 9452–9472 (1986).

Calibration of the magnetic compass of a migratory bird by celestial rotation

Kenneth P. Able & Mary A. Able

Department of Biological Sciences, State University of New York at Albany, New York 12222, USA

MIGRATORY birds use a variety of environmental stimuli in orientation. Species that migrate primarily at night can use compasses based on the geomagnetic field, stars, the Sun and patterns of skylight polarization^{1–4}. These compass mechanisms can interact both when migratory birds make day-to-day orientation decisions⁵ and during their ontogeny in young birds^{6,7}. All of the known compasses used by migratory birds seem to be modifiable by experience during early development. For example, a functional magnetic compass develops in birds that have never seen the sky^{8–11}. But the preferred direction of orientation by the magnetic compass may be modified during the first three months of life by exposing naive birds to the sky under conditions in which magnetic directions differ substantially from compass directions indicated by the Sun and stars (true or geographic directions)^{6,9,10}. For hand-raised Savannah sparrows (*Passerculus sandwichensis*), experience with either the clear daytime or night sky is sufficient to effect this calibration of the magnetic compass¹¹. We therefore proposed that celestial rotation, which provides a source of geographic directions both day and night, is the calibrating reference. Here we report that the rotation of an artificial pattern of 'stars' calibrates the preferred direction of magnetic orientation of young Savannah sparrows.

The Savannah sparrow is a medium-distance nocturnal migrant that breeds in grassland, meadow and tundra across North America from the northern United States northwards to 60–70° N. It migrates over winter to the southern United States southwards to northern Central America. Adult Savannah sparrows possess magnetic^{11,12} and star compasses⁹, and use visual cues associated with clear sunset skies (including skylight polarization patterns)^{13–15}. To examine the ontogeny of magnetic orientation, we transferred nestling sparrows from nests in the field to a laboratory incubator before their eyes opened (2–3 days post-hatching) and hand-raised them to independence. During this period, the birds lived in a normal magnetic field

in the ambient photoperiod, but had no view of the outdoors or sky. At 30–40 days old, 20 birds from six nests were each assigned to one of four groups (N, S, E, W) such that nestmates were, to the greatest extent possible, placed in different groups. Nightly, the four groups of birds were placed in box cages at the four cardinal positions under an artificial sky with an arbitrary pattern of 16 'stars' (Fig. 1)¹⁶, where they remained

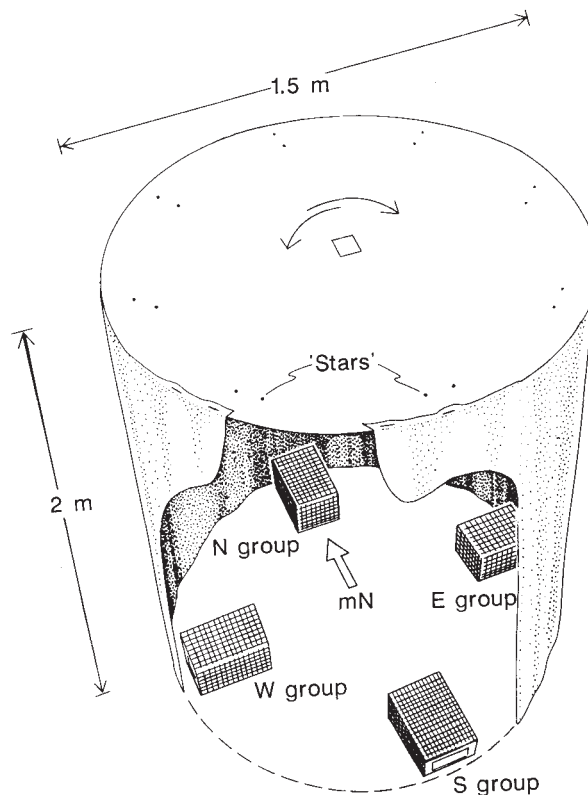


FIG. 1 Diagram of the experimental apparatus. The artificial 'sky' was an exact replica of that described by Wiltschko and Wiltschko¹⁶. A shroud of black plastic enclosed the rotating disk with its 16 'stars' and the cages in which the four groups of birds spent the night. Note the directional relationships between the centre of rotation of the 'sky' and the magnetic directions experienced by the birds in each of the four groups.

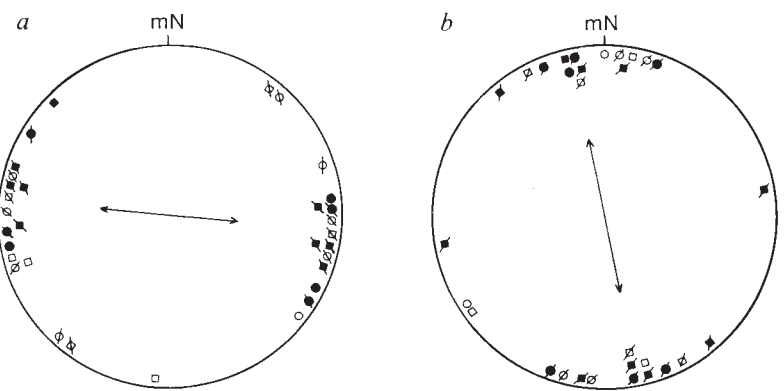
FIG. 2 Magnetic orientation of the four groups of birds tested indoors in orientation cages covered with white translucent plastic sheets. *a*, Orientation of the E (circles) and W (squares) groups. Solid symbols refer to tests in the unshifted magnetic field, open symbols to tests of the same birds in a field in which magnetic north was shifted 90° clockwise. Because the directions chosen were the same under the two field conditions, the data were pooled for statistical analysis with a single datum included for each bird. These groups showed east-west magnetic orientation ($r=0.791$, modes = 96° and 276°, $P<0.001$, Rayleigh test, $n=10$ birds). *b*, Orientation of the N (circles) and S (squares) groups. These groups showed north-south magnetic orientation ($r=0.915$, modes = 349° and 169°, $P<0.001$, Rayleigh test, $n=9$ birds). These two distributions are significantly different ($P<0.001$, Watson's U^2 test). Points with a slash mark indicate hopping activity of birds that showed statistically bimodal orientation²⁰. Magnetic north, mN.

METHODS. Orientation tests were performed in a wooden frame garage with an ambient magnetic field of total intensity = 5.3×10^4 nT, inclination = 69°. The shifted field was produced with Helmholtz coils (2-m diameter) and had a total intensity of 5.4×10^4 nT, inclination = 70°. Birds were prevented from seeing outside the test cages by the translucent covers. A dim fluorescent light, passing through two diffusers and the translucent cage cover, provided a light of 0.2–0.4 lux in the cages. This variation in light

all night. The sky disk rotated around its centre at a uniform rate of 15° h^{-1} ; the sense of rotation (clockwise or counterclockwise) and starting position of the disk were changed randomly from night to night. The birds remained in the cages all night for a total of 22 nights of experience. When the birds came into migratory condition for the first time in late September, their directions of magnetic orientation were recorded in covered Emlen funnel orientation cages¹⁷ during tests conducted indoors in normal and shifted magnetic fields.

If the centre of rotation of the artificial starry sky defines geographic north for the birds, then each group of birds would be expected to perceive geographic north in a different magnetic direction (Fig. 1). So if the preferred directions of magnetic orientation of young birds are modifiable by geographic information, then these groups of birds should differ in their magnetic orientation. For example, the birds of the N group, viewing the artificial sky from an eccentric position towards magnetic north from the centre of rotation, would perceive the centre of rotation, and therefore geographic north, to be towards magnetic south. Hand-raised Savannah sparrows typically show axially bimodal migratory orientation when tested over several nights^{9,11}. Individual birds usually orient in a unimodal direction in each test, but often switch to more or less opposite directions on subsequent nights. This being the case, the four groups in this experiment become two groups: (1) the E and W groups predicted to orient magnetic east-west (the axis that corresponded to north-south as defined by the stellar rotation observed before the tests); and (2) the N and S groups predicted to orient magnetic north-south.

Tested in the ambient geomagnetic field and in a field of ambient intensity and inclination, but with magnetic north shifted 90° clockwise, the birds oriented as shown in Fig. 2. The birds selected the same magnetic directions in the shifted and unshifted field conditions, indicating that the observed orientation was in fact based on the magnetic stimulus. The orientation of all four groups was axial, as expected, and there was no suggestion of polarity in any group. The E and W groups oriented on a magnetic east-west axis (96°–276°) (Fig. 2*a*), whereas the N and S groups oriented on a magnetic north-south axis (349°–169°) (Fig. 2*b*). These two distributions are significantly different and the angular difference between them (73°) does not differ from the expected and maximum possible difference of 90°. Control birds raised indoors in a normal magnetic field and without stellar rotation oriented on a magnetic northwest-southeast axis (length of mean vector $r=0.710$, modes =



intensity occurred across the array of test cages; light intensity in each cage was uniform. Tests began within 1 h after sunset and lasted for 3 h. Roughly half the birds were first tested in the normal field condition, the remainder first tested in the shifted field. Thereafter, individuals were assigned randomly to one or the other test condition. Each point represents the mean of the modal orientation directions of each bird (each bird tested 5–9 times). The orientation records were analysed blind with respect to treatment group and test condition²¹.

335°–155°; $P<0.01$, Rayleigh test), statistically indistinguishable from the N and S groups¹¹.

The results of this experiment show that exposure to the axis of rotation of an artificial star pattern during the first three months of life can alter the preferred magnetic direction for migratory orientation. Celestial rotation, which may be assessed at night from the stars or during daytime from the Sun's path or patterns of skylight polarization, provides information about the location of true north, and the axis of stellar rotation defines directions during the ontogeny of the star compass in night migrants^{18,19}. Whether, and if so, how, celestial rotation serves as the calibrating reference for the magnetic compass during daytime remains to be determined.

Migratory birds seem to possess two innate representations of the migration direction, one relative to celestial rotation and another relative to the magnetic field⁵. The development of a functional star compass requires only exposure to stellar rotation and is shielded during early development from any magnetic influence. Adult migrants of some species can recalibrate their star compass according to magnetic cues⁵. Whereas a magnetic orientation capability develops in birds that have never seen the stars or Sun, it is modifiable early in life but is apparently rigid later when the birds are migrating. This developmental flexibility in magnetic orientation may have evolved to enable birds to cope with spatial variation in the geomagnetic field, especially variation in magnetic declination across the breeding range. The capability to adjust magnetic orientation by a reference of geographic compass directions, which are the relevant directions to a migratory bird moving over the earth, may be important especially to high latitude breeders who often experience a large magnetic declination. How a bird copes with the spatial variability in declination encountered later during migration remains unclear. □

Received 24 March; accepted 18 July 1990.

1. Wiltschko, W. *Comp. Biochem. Physiol.* **76A** 709–717 (1983).
2. Baker, R. R. *Bird Navigation: The Solution of a Mystery?* (Holmes & Meier, London, 1984).
3. Able, K. P. & Cherry, J. D. in *Migration: Mechanisms and Adaptive Significance* (ed. Rankin, M. A.) 516–525 (Marine Sci. Inst., Univ. Texas, Port Aransas, 1985).
4. Moore, F. R. *Biol. Rev.* **62**, 65–86 (1987).
5. Wiltschko, W. & Wiltschko, R. *Trends Ecol. Evol.* **3**, 13–15 (1988).
6. Able, K. P. & Bingman, V. P. *Q. Rev. Biol.* **62**, 1–29 (1987).
7. Able, K. P. *Experientia* **46**, 388–394 (1990).
8. Wiltschko, W. & Gwinner, E. *Naturwissenschaften* **61**, 406 (1974).
9. Bingman, V. P. *Behaviour* **87**, 43–53 (1983).
10. Bingman, V. P., Beck, W. & Wiltschko, W. in *Migration: Mechanisms and Adaptive Significance* (ed. Rankin, M. A.) 544–552 (Marine Sci. Inst., Univ. Texas, Port Aransas, 1985).
11. Able, K. P. & Able, M. A. *Anim. Behav.* **39**, 905 (1990).

12. Bingman, V. P. *Anim. Behav.* **29**, 962-963 (1981).
13. Moore, F. R. *Anim. Behav.* **28**, 684-704 (1980).
14. Moore, F. R. *Anim. Behav.* **33**, 657-663 (1985).
15. Able, K. P. & Able, M. A. *Anim. Behav.* **39**, 1189 (1990).
16. Wiltchko, W. & Wiltchko, R. *J. comp. Physiol.* **A109**, 91-99 (1976).
17. Emlen, S. T. & Emlen, J. T. *Auk* **83**, 361-367 (1966).
18. Emlen, S. T. *Science* **170**, 1198-1201 (1970).
19. Wiltchko, W., Daum, P., Fergenbauer-Kimmel, A. & Wiltchko, R. *Ethology* **74**, 285-292 (1987).
20. Batschelet, E. *Circular Statistics in Biology* (Academic, New York, 1981).
21. Cherry, J. D. & Able, K. P. *Auk* **103**, 225-227 (1986).

ACKNOWLEDGEMENTS. We thank C. Walcott for use of the magnetometer, and R. L. Holberton for comments on the manuscript. This work was supported by the NSF.

Genetic loads and estimates of mutation rates in highly inbred plant populations

B. Charlesworth, D. Charlesworth & M. T. Morgan

Department of Ecology and Evolution, University of Chicago,
5630 S. Ingleside Avenue, Chicago, Illinois 60637, USA

ESTIMATES of the rate of mutation to mildly deleterious alleles are fundamental to the testing of theories of the maintenance of sexual reproduction¹ and recombination^{2,3}. There are, however, few estimates of this parameter^{4,5}. We suggest that data on inbreeding depression in highly self-fertilizing populations might provide such estimates. The effect of inbreeding in reducing genetic load and inbreeding depression is well known^{6,7}, but little attention has been paid to the magnitude of these quantities in highly inbred natural populations. Modelling of mutational load shows that highly self-fertilizing populations should still have some genetic load and inbreeding depression⁸⁻¹⁰. In such populations, inbreeding depression is the converse of heterosis and its magnitude can be estimated from the increase in fitness-related characteristics in individuals produced by intercrossing different strains. Here we derive the inbreeding depression expected in completely inbred populations, and use it to estimate mutation rates per genome for some plant species.

Three parameters have important influences on the magnitudes of the genetic load and inbreeding depression due to deleterious mutations (Table 1). These are the dominance coefficient, h , the rate of self-fertilization of the population, S , and the mutation rate, U (the mean number of new mutant alleles arising per generation per diploid genome). Higher mutation rates lead to greater genetic load and inbreeding depression (Table 2), as do lower dominance coefficients, whereas higher selfing rates reduce them^{6,8,9}. The selection coefficient s against homozygous mutations (defined in Table 1), however, has only a slight influence on the genetic load and inbreeding depression, although it strongly affects the numbers of mutations per individual^{6,8,9}.

For complete selfing (all loci homozygous), approximate analytical results can be obtained, assuming that the fitness of a genotype is the product of the fitnesses at each of its loci (multiplicative fitnesses). The distribution of the number of (homozygous) mutations per individual then follows a Poisson distribution with mean $\bar{n}$ (refs 10, 11). The mean fitness of progeny produced by selfing (which is also the mean fitness of a selfing population) is thus

$$\bar{w}_s \approx e^{-\bar{n}} \left\{ 1 + (1-s)\bar{n} + \frac{(1-s)^2 \bar{n}^2}{2!} + \dots \right\}$$

$$= e^{-\bar{n}} \cdot e^{\bar{n}(1-s)} = e^{-s\bar{n}}$$

Furthermore, the mean fitness under complete selfing can be shown to be $\sim \exp(-U/2s)$, regardless of the selection model¹⁰. Hence $\bar{n} \approx U/2s$ (see also refs 11, 12). The fitness of the progeny of a cross between two lines homozygous for n_i and n_j mutations is $(1-hs)^{n_i+n_j}$, where $n_i + n_j$ is Poisson distributed with mean

$2\bar{n} \approx U/s$. Therefore the mean fitness of the progeny of crosses between different individuals is

$$\bar{w}_x \approx \exp(-hs \cdot U/s) = \exp(-hU)$$

and the inbreeding depression is

$$\delta = 1 - \frac{\bar{w}_s}{\bar{w}_x} \approx 1 - \exp\left[-\left(\frac{1}{2} - h\right)U\right]$$

and thus

$$U \approx \frac{-2 \ln(1-\delta)}{1-2h} \quad (1)$$

Some values based on this calculation are shown in Table 2. Exact computer calculations give excellent agreement, with selfing rates from 0.9-0.99 (Table 2). For the biologically more plausible case of synergistic interactions between loci (when fitness decreases more rapidly with increasing numbers of mutations in the genotype^{1,2,6,10}), the results are similar, for weak synergism. With strong synergism, the inbreeding depression is higher than for the multiplicative case, especially with high dominance (Fig. 1), but the selection model has only a minor effect on the inbreeding depression values calculated for these highly inbred populations.

The mutational model for populations with high selfing rates can thus yield large inbreeding depression values, given sufficiently high mutation rates and low dominance coefficients per genome. The values are not very sensitive to the selection

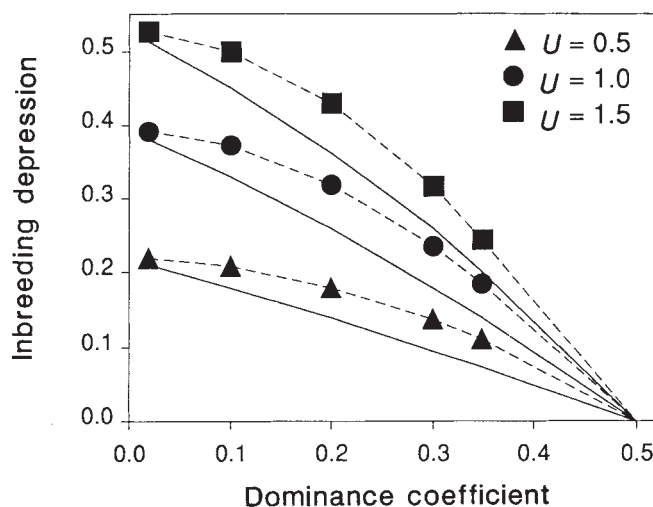


FIG. 1 Inbreeding depression values under the mutational load model with synergistic fitness interactions and with various mutation rates, in equilibrium populations with selfing rate 0.99 (solid symbols). The solid lines show the values calculated by the approximation described in the text, assuming multiplicative fitness interactions. The dashed lines refer to the results of the computer calculations of the synergistic model of fitness interactions. This model is a generalization of the quadratic fitness model of Kimura and Maruyama²⁹, in which the fitness of a genotype with n mutations is given by $w_n = 1 - an - bn^2$ (ref. 6). To take into account heterozygous and homozygous mutations, n in this expression can be expressed as an effective number of mutations, by weighting heterozygous mutations by the dominance coefficient, so that one homozygous mutation is equivalent to $1/h$ heterozygous ones³⁰; n is then the sum of the number of homozygous mutations and the weighted number of heterozygous mutations. We used this n in the fitness function $w_n = \exp[-(\alpha n + (\beta n^2)/2)]$. The coefficient α in this epistatic fitness expression can be viewed as measuring the strength of selection, and β/α as measuring the degree of synergism. These coefficients are closely related to the parameters in the quadratic fitness model because, with weak selection, $\alpha \approx a$ and $\beta \approx 2b$. The parameter values used here were $\alpha = 0.1$, $\beta = 0.2$, which represents very strong synergism¹⁰.

TABLE 1 Model of mutational load due to mutations at many unlinked loci

Genotype	AA	Aa	aa
Fitness	1	1-hs	1-s

The relative fitnesses of different genotypes at one of the loci controlling fitness are shown. The parameter s is the selection coefficient against mutant homozygotes, and h is the dominance coefficient which describes the extent to which heterozygotes express the harmful effects of the mutant alleles⁶. To carry out the calculations deterministically, without stochastic noise, we used the method of Kondrashov¹, which assumes that all mutations have the same effects on fitness, and occur at unlinked loci. The number of new mutations per zygote is assumed to follow a Poisson distribution, with mean U . The state of a population is then completely described by the distribution of the numbers of loci homozygous and heterozygous for mutant alleles. The mean fitness of progeny produced by outcrossing is denoted by $\bar{w}_x$, and that of progeny of self-fertilization by $\bar{w}_s$. The genetic load ($1 -$ the mean fitness of individuals in the population) in a population whose selfing rate is S is thus $1 - S\bar{w}_s - (1 - S)\bar{w}_x$, and the inbreeding depression of the population⁷ is defined as $\delta = 1 - \bar{w}_s/\bar{w}_x$.

model, or to the strength of selection. These results support the use of equation (1) to estimate the mutation rate per genome. The rate estimated will be for mildly deleterious alleles, because it is unlikely that highly inbred populations will carry appreciable frequencies of recessive lethal mutations⁷. They are also unlikely to be polymorphic for overdominant alleles at many loci, as such polymorphisms cannot be maintained in highly selfing populations unless the selection coefficients against the two homozygotes are nearly equal, which is improbable¹³. Furthermore, the weight of the evidence indicates that overdominant loci do not contribute much to genetic load in natural populations^{8,14}.

To use equation (1) to estimate U , an estimate of h is needed. This parameter has rarely been estimated for mildly detrimental mutations in organisms other than *Drosophila*, although estimates for lethal mutations in plant populations exist, and suggest low dominance coefficients for such alleles¹⁵. For mildly detrimental alleles, estimates could be obtained by studying the results of crosses between inbreeding populations, from the regression of hybrid fitness values on the mid-parental pure-line values⁴. In the absence of estimates of h , it is conservative to assume low values. The data on inbreeding depression then given rough estimates of U .

Table 3 shows the few available inbreeding depression data published for very inbred plant species. The data suggest that

TABLE 2 Theoretical calculations of inbreeding depression values in highly selfing populations

Mutation rate U	Dominance coefficient h	Analytical inbreeding depression values ($S=1.0$)	Inbreeding depression values from computer runs		
			$S=0.99$	$S=0.95$	$S=0.9$
0.5	0.02	0.213	0.215	0.224	0.236
	0.1	0.181	0.179	0.185	0.193
	0.2	0.139	0.135	0.138	0.142
	0.3	0.095	0.090	0.091	0.094
	0.35	0.072	0.068	0.069	0.070
1.0	0.02	0.381	0.384	0.400	0.424
	0.1	0.330	0.327	0.337	0.352
	0.2	0.259	0.252	0.257	0.264
	0.3	0.181	0.172	0.174	0.177
	0.35	0.139	0.131	0.132	0.134
1.5	0.02	0.513	0.517	0.539	0.574
	0.1	0.451	0.448	0.461	0.481
	0.2	0.362	0.353	0.359	0.368
	0.3	0.259	0.247	0.249	0.253
	0.35	0.201	0.190	0.191	0.192

The approximate analytical results derived here for complete selfing are shown, as well as the results calculated by the computer runs with multiplicative fitness interactions for three different selfing rates (S), with selection coefficient $s=0.2$.

heterosis in highly inbreeding plants is large enough that the value of U must be at least 0.5, even assuming low dominance, and possibly double that if h were higher (Tables 2 and 3). The assumption of multiplicative fitness interactions in the estimation equation leads to overestimation of U , because synergistic interactions tend to produce larger inbreeding depression values in highly selfing populations, but the effect of even strong synergism is slight (Fig. 1). On the other hand, the values in Table 3 are all based on single components of fitness. If other fitness components that were not measured also showed heterosis, the values shown in the table would be underestimates of the total effect of mutations on fitness, and the true values of U would be higher. Linkage also decreases the inbreeding depression caused by a given mutation rate, compared with the results obtained in our models which assumed unlinked loci^{9,10}, but the effects should be slight in highly selfing populations,

TABLE 3 Estimates of inbreeding depression values for several highly inbreeding plant species

Species	Fitness component	Inbreeding depression estimate ($1 - \bar{w}_s/\bar{w}_x$)	Breeding system	References
<i>Arabidopsis thaliana</i>	Yield (=plant dry weight)	0.07*–0.16†	Highly selfing	20–23
<i>Thlaspi arvense</i>	Germination rate	0.17‡	Highly selfing	24
<i>Impatiens capensis</i>	Yield	0.155–0.265§	$S=0.95$	25, 26
<i>Avena fatua</i>	Tiller number	0.58	Highly selfing	27
	Survival to flowering	0.47	Highly selfing	27
<i>Scleranthus annuus</i>	Stamen fertility¶	0.18	Highly selfing	28

* Based on data from 25 °C, the temperature at which all plants, whether inbred or outcrossed, grew best. The inbreeding depression values for all six temperatures used ranged from 0.037 to 0.672.

† Probably an underestimate, because this value is based on the comparison of parental strains with the F_2 -generation between them, due to low yields of F_1 seeds.

‡ Based on mean values from three experiments, treating inter- and intrapopulation crosses equivalently. The estimates depend on assumptions about the selfing rate of the population.

§ Based on adjusted weights of plants from seeds of cleistogamous (selfing) and chasmogamous (open-pollinated) flowers, and on plants grown in various competitive situations.

|| The published data are for selfed and open-pollinated flowers. We assumed an outcrossing rate of 0.1 for open-pollinated flowers. Higher inbreeding depression values would be obtained if the outcrossing rate were lower than this. The values from wild oats are inconsistent with the other values, and may be suspect because the two types of seeds were not obtained from flowers treated in the same way. It is possible that the bagging procedure used to ensure that outcrossing did not occur could have lowered the fitness values of the selfed progeny.

¶ Based on the five stamens with high fertility.

which will be nearly completely homozygous. The mutation rate estimates are similar to those deduced from experiments on *Drosophila*^{4,16}, and are much higher than lethal mutation rates in plants¹⁷. These data, although scanty, show that highly selfing populations can have measurable inbreeding depression, as the mutational model predicts, and are consistent with data from highly inbreeding crops^{18,19}.

The considerations outlined here suggest that further studies of inbreeding depression in highly inbred populations would be worthwhile. It will also be important to extend these to include independent estimation of the mutation rate to detrimental alleles. This could be done by experiments in which mutations are allowed to accumulate, as in the work of Mukai in *Drosophila*¹⁶. □

Received 14 May; accepted 31 July 1990.

1. Kondrashov, A. S. *Genetics* **111**, 635–653 (1985).
2. Kondrashov, A. S. *Genet. Res.* **44**, 199–217 (1984).
3. Charlesworth, B. *Genet. Res.* **55**, 199–222 (1990).
4. Crow, J. F. & Simmons, M. J. in *The Genetics and Biology of Drosophila* (eds Ashburner, M., Carson, H. L. & Thompson, J. N.) 1–35 (Academic, London, 1983).
5. Kondrashov, A. S. *Nature* **336**, 435–440 (1988).
6. Crow, J. F. in *Mathematical Models in Population Genetics* (ed. Kojima, K.-I.) 128–177 (Springer, Berlin, 1970).

7. Lande, R. & Schemske, D. W. *Evolution* **39**, 24–40 (1985).
8. Charlesworth, D. & Charlesworth, B. *A. Rev. Ecol. Syst.* **18**, 237–268 (1987).
9. Charlesworth, D., Morgan, M. T. & Charlesworth, B. *Evolution* (in the press).
10. Charlesworth, B., Morgan, M. T. & Charlesworth, D. *Genet. Res.* (in the press).
11. Heller, J. & Maynard Smith, J. *Genet. Res.* **32**, 289–293 (1979).
12. Hopf, F. A., Michod, R. E. & Sanderson, M. J. *Theoret. Populat. Biol.* **33**, 243–265 (1988).
13. Kimura, M. & Ohta, T. *Theoretical Topics in Population Genetics* (Princeton University Press, Princeton, New Jersey, 1971).
14. Houle, D. *Genetics* **123**, 789–801 (1989).
15. Ohnishi, O. *Japanese J. Genet.* **57**, 623–639 (1982).
16. Mukai, T. in *Population Genetics and Molecular Evolution* (eds Ohta, T. & Aoki, K.-I.) 125–145 (Springer, Berlin, 1985).
17. Klekowski, E. J. & Godfrey, P. J. *Nature* **340**, 389–391 (1989).
18. Matzinger, D. F. in *Statistical Genetics and Plant Breeding* (eds Hanson, W. D. & Robinson, H. F.) 253–279 (NAS-NRC, Washington, DC, 1963).
19. Wright, S. *Evolution and the Genetics of Populations, Vol. 3. Experimental Results and Evolutionary Deductions* (Univ. Chicago Press, Chicago, 1977).
20. Griffing, B. & Langridge, J. in *Statistical Genetics and Plant Breeding* (eds Hanson, W. D. & Robinson, H. F.) 368–394 (NAS-NRC, Washington DC 1963).
21. Griffing, B. *Genetics* **122**, 943–956 (1989).
22. Abbott, R. J. & Gomes, M. F. *Heredity* **62**, 411–418 (1988).
23. Snape, J. W. & Lawrence, M. J. *Heredity* **27**, 299–302 (1971).
24. Riley, R. *New Phytol.* **55**, 319–330 (1956).
25. Schmitt, J. & Erhardt, D. W. *Evolution* **44**, 269–278 (1990).
26. Schmitt, J., Eccleston, J. & Ehrhardt, D. W. *Oecologia* **72**, 341–347 (1987).
27. Imam, A. G. & Allard, R. W. *Genetics* **51**, 49–62 (1965).
28. Svensson, L. *Evol. Trends Plants* **2**, 31–39 (1988).
29. Kimura, M. & Maruyama, T. *Genetics* **54**, 1337–1351 (1966).
30. Sved, J. & Wilton, A. N. *Genet. Res.* **54**, 119–128 (1989).

ACKNOWLEDGEMENTS. This work was supported by the NSF.

Expression and characterization of the cystic fibrosis transmembrane conductance regulator

Richard J. Gregory, Seng H. Cheng, Devra P. Rich*, John Marshall, Sucharita Paul, Kathleen Hehir, Lynda Ostedgaard*, Katherine W. Klinger†, Michael J. Welsh* & Alan E. Smith‡

Genzyme Corporation and † IG Laboratories Inc., One Mountain Road, Framingham, Massachusetts 01701, USA

* Howard Hughes Medical Institute, Departments of Internal Medicine and Physiology and Biophysics, University of Iowa, College of Medicine, Iowa City, Iowa, USA

CYSTIC fibrosis (CF) is a common lethal genetic disease that manifests itself in airway and other epithelial cells as defective chloride ion absorption and secretion^{1,2}, resulting at least in part from a defect in a cyclic AMP-regulated, outwardly-rectifying Cl[−] channel in the apical surface^{3–5}. The gene responsible for CF has been identified and predicted to encode a membrane protein termed the CF transmembrane conductance regulator (CFTR)^{6–8}. Identification of a cryptic bacterial promoter within the CFTR coding sequence led us to construct a complementary DNA in a low-copy-number plasmid, thereby avoiding the deleterious effects of CFTR expression on *Escherichia coli*. We have used this cDNA to express CFTR *in vitro* and *in vivo*. Here we demonstrate that CFTR is a membrane-associated glycoprotein that can be phosphorylated *in vitro* by cAMP-dependent protein kinase. Polyclonal and monoclonal antibodies directed against distinct domains of the protein immunoprecipitated recombinant CFTR as well as the endogenous CFTR in nonrecombinant T84 cells. Partial proteolysis fingerprinting showed that the recombinant and non-recombinant proteins are indistinguishable. These data, which establish several characteristics of the protein responsible for CF, will now enable CFTR function to be studied and will provide a basis for diagnosis and therapy.

Initial attempts to assemble a full-length CFTR coding sequence resulted in extensive rearrangements in the DNA

clones obtained. This, together with the reported absence of full-length clones in the original isolates, made us consider whether transfected bacteria might be selecting against the full-length CFTR cDNA. Although the CFTR DNA sequence includes some repeated elements, they are not more frequent or extensive than is typical for other stable DNA sequences. This argued against intrinsic DNA instability as the cause of the problem but instead suggested that the presence of the full-length cDNA was toxic to cells. Such toxicity generally results from the inappropriate expression of cDNA and could be due to the presence of cryptic promoters within the DNA sequence. To search for any such transcription initiation sites, segments of CFTR were placed upstream of a promoterless gene encoding chloramphenicol acetyl transferase (CAT), transfected into *E. coli*, and the bacteria challenged with chloramphenicol⁹. By this means we identified an active promoter sequence beginning at residue 908 in exon 6B (Fig. 1a). This sequence shows good homology with the consensus *E. coli* promoter sequence, having 9 of 13 residues identical, including a highly conserved T residue at the 3' end of the −10 box (Fig. 1b)¹⁰.

To circumvent this problem of cDNA toxicity, we reduced the gene dosage of the cDNA by assembling a full-length CFTR open reading frame in a vector containing a low copy number of DNA replication originally derived from the plasmid pSC101 (ref. 11) (pSC-CFTR-2, Fig. 1c). A second plasmid pSC-CFTR-2ΔF508 was constructed which contains a deletion corresponding to amino-acid residue 508, the site of a common CF mutation⁷. The resulting clones had no deleterious effect on bacterial growth and were stable. DNA sequence analysis of the complete coding sequence revealed a sequence identical to that reported by Riordan *et al.*⁷, with the exception that the base at position 1,990 is C rather than A.

To examine the protein product of the CFTR cDNA, we constructed low-copy-number vectors (pTM-CFTR-3 and pTM-CFTR-3ΔF508) with the protein coding sequence immediately downstream of a bacteriophage T7 promoter and an encephalomyocarditis (EMC) virus internal ribosome entry sequence and followed by a T7 transcriptional terminator^{12,13}. The T7 and EMC sequences were isolated from the plasmid PTM-1 (T. Mizukami, O. Elroy-Stein and B. Moss, personal communication). After transcription *in vitro* using T7 RNA polymerase, the resulting RNA was translated in a reticulocyte lysate in the presence of [³⁵S]methionine. Figure 2a shows an autoradiograph of the cell-free product after gel electrophoresis. A major band migrating with a relative molecular mass of

‡ To whom correspondence should be addressed.

130,000 (M_r 130K) was detected (band A, lane 2). When a truncated transcript generated by linearization at the Tth1111 site (amino acid 1,336) was translated, the M_r of the cell-free product (lane 5) was, as expected, reduced by ~10 K. The M_r of the *in vitro* product is somewhat smaller than that expected for a protein consisting of 1,480 amino acids. But Greenberger *et al.*¹⁴ have reported that the related multiple drug resistance protein also migrates faster than expected on polyacrylamide gels.

In vitro translation was enhanced by the presence of detergent (0.5% Triton X-100)¹⁵ in the reticulocyte cell-free extract (lane 6). Furthermore, boiling the protein in sample buffer before electrophoresis dramatically reduced the amount of material migrating at 130 K and increased the material at the top of the gel (lane 8). Both these observations are reminiscent of results¹⁴⁻¹⁶ with other integral membrane proteins and are probably a consequence of the presence of several hydrophobic transmembrane sequences within the polypeptide.

When translation in the reticulocyte lysate was performed in the presence of canine microsomal membranes, the mobility of the cell-free product decreased to correspond to M_r ~135 K (band B, lane 3). This implies that band A is modified by the membranes by addition of core carbohydrate. Consistent with this, when band B was treated with detergent and *N*-Glycanase its mobility increased to comigrate with band A (lane 4). Translation of the $\Delta F508$ mutant cDNA gives a primary *in vitro* product indistinguishable from wild type (lane 7).

To examine the expression of CFTR in transfected cells we used the vaccinia virus/T7 RNA polymerase hybrid expression system^{13,17}. HeLa cells were infected with a vaccinia virus encoding the RNA polymerase of bacteriophage T7 and subsequently transfected with the vector pTM-CFTR-3 or control plasmids. Figure 2b shows a polyacrylamide gel of the proteins labelled with [³⁵S]methionine after synthesis in infected and uninfected cells. Material co-migrating with bands A and B is present in infected cells transfected with the pTM-CFTR-3 vector, but not in those transfected with the control plasmid pTM-1 or in uninfected cells. Treatment with *N*-Glycanase resulted in a mobility shift of the upper of the two bands to generate material comigrating with band A (data not shown). Although not visible in Fig. 3, another band of M_r 150 K is also present in small amounts in the cells transfected with CFTR cDNA (see later).

To characterize further the CFTR gene product we raised antisera in rabbits against fusion proteins containing the amino-acid sequences encoded by exon 13 (polyclonal antibody Ex13) or exon 10 (polyclonal antibody Ex10) fused to β -galactosidase. Exon 13 encodes the unique, polar, R domain⁷. Exon 10 encodes a portion of the first nucleotide-binding fold, including the site of the common CF-related deletion, $\Delta F508$ ⁷. Mouse monoclonal antibodies to the exon 13 fusion protein (monoclonal antibodies 13-1 and 13-2) were also produced, screened and cloned using standard procedures¹⁸.

Figure 3a shows the immunoprecipitation of the cell-free product made in the presence of membranes and of the protein

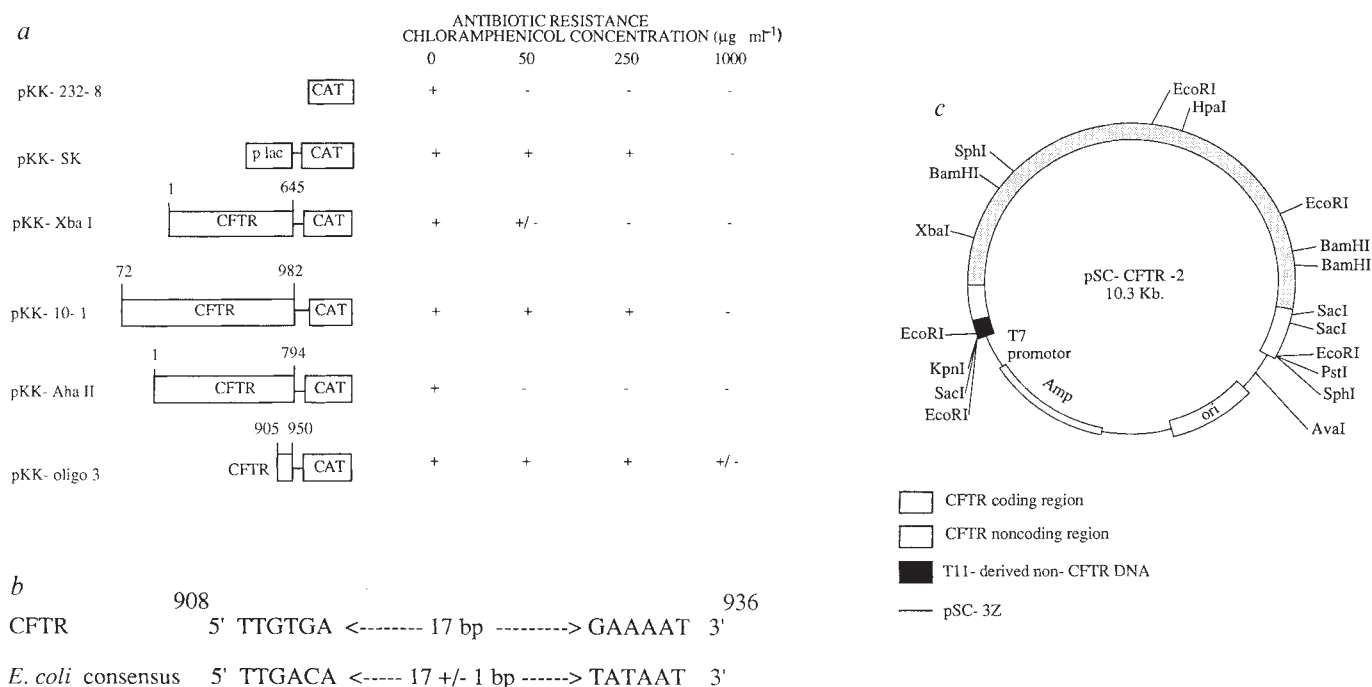


FIG. 1 Identification of a functional *E. coli* promoter within the CFTR cDNA and assembly of the complete CFTR open reading frame. **a**, Plasmid clones used to identify promoter sequences within the CFTR cDNA by activation of a promoterless CAT gene¹⁰. *E. coli* cells containing the constructs were challenged with increasing concentrations of chloramphenicol. Control plasmids contained either no insert (pKK232-8) or the β -galactosidase promoter (p lac) from pBluescript SK⁻ (Stratagene). Nucleotide coordinates for the CFTR gene fragments are as given by Riordan *et al.*⁷. Symbols: +, growth to confluence; +/-, slow growth; -, no growth. **b**, Alignment of the promoter sequence within the CFTR cDNA with the consensus *E. coli* promoter sequence¹¹. **c**, Map of the low-copy-number plasmid pSC-CFTR-2, which contains the first 5.5 kilobases of the CFTR sequence of Riordan *et al.*⁷, including the entire open reading frame.

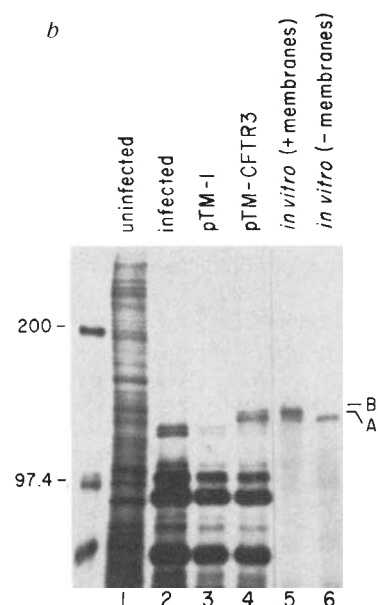
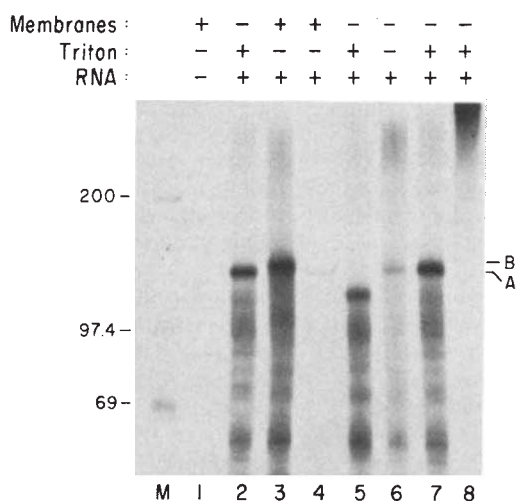
METHODS. DNA manipulations were according to published procedures²⁶.

All CFTR plasmids were constructed using the partial CFTR cDNA clones T11, T16-1, T16-4.5 C1-1/5 and 10-1 (ref. 7), (obtained from the American Type Culture Collection), except for PKK-oligo3, which contains synthetic oligonucleotides corresponding to nucleotides 905-950 of the CFTR cDNA. The CFTR open reading frame was reconstructed in a plasmid (pSC-3Z) containing the pSC101 origin of replication from plasmid pLG338^{11,27} combined with the β -lactamase gene and polylinker of pGEM-3Z (Promega). Base pairs (bp) 1-645 of pSC-CFTR-2 were derived from T16-1, bp 646-1,817 from T16-1 and bp 1,818-5,576 from T16-4.5. The CFTR open reading frame within pSC-CFTR-2 was completely sequenced by the dideoxy sequencing method. To construct a $\Delta F508$ variant of pSC-CFTR-2 (pSC-CFTR-2 $\Delta F508$), the sequence between nucleotide positions 1,504 and 2,463 was replaced with the corresponding segment from C1-1/5, which contains the $\Delta F508$ mutation.

FIG. 2 *In vitro* and *in vivo* expression of the CFTR protein. *a*, CFTR cDNA transcripts were synthesized from plasmids pTM-CFTR-3 and pTM-CFTR-3ΔF508 using T7 RNA polymerase and translated in a rabbit reticulocyte lysate cell-free system. Translation products were analysed by electrophoresis on 6% SDS-polyacrylamide gels²⁸ and fluorography. Lane 1, control (no RNA added); lane 2, pTM-CFTR-3 transcript plus Triton X-100; lane 3, pTM-CFTR-3 transcript plus microsomal membranes; lane 4, as in lane 3 then treated with *N*-Glycanase²⁹; lane 5, pTM-CFTR-3 transcript truncated after codon 1,336, plus Triton X-100; lane 6, pTM-CFTR-transcript translated without Triton X-100 or microsomal membranes; lane 7, pTM-CFTR-3ΔF508 transcript plus Triton X-100; lane 8, as in lane 2 except boiled for 3 min before electrophoresis. *b*, Expression of CFTR in transfected HeLa cells. CFTR expression was directed by a vaccinia virus/bacteriophage T7 hybrid expression system where transcription

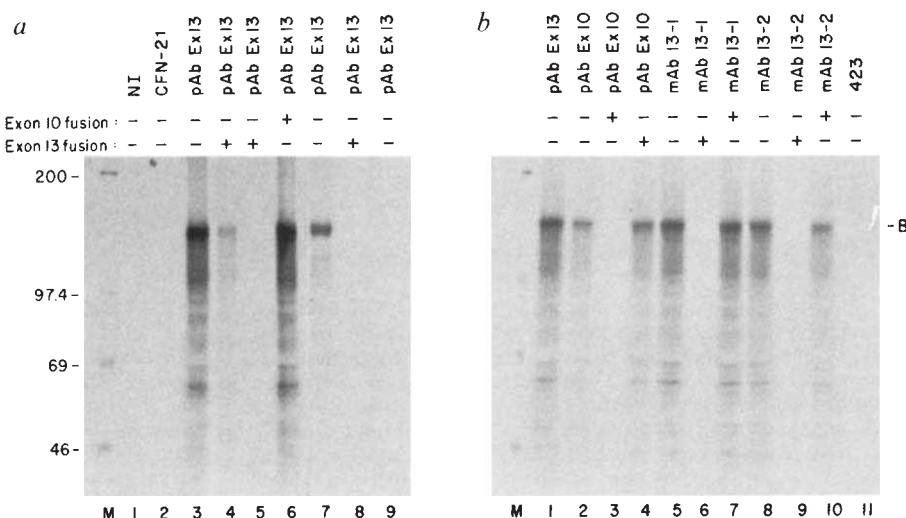
of transfected genes is catalysed by T7 RNA polymerase provided *in trans* by a recombinant vaccinia virus vector^{13,17}. Cell lysates were prepared from HeLa cells labelled with [³⁵S]methionine and analysed as above. Lane 1, uninfected HeLa cells; lane 2, HeLa cells infected with vaccinia virus; lane 3, vaccinia-infected cells transfected with the control plasmid pTM1; lane 4, vaccinia-infected cells transfected with pTM-CFTR-3; lane 5, pTM-CFTR-3 *in vitro* translation product prepared in the presence of microsomal membranes; lane 6, as in lane 5 but prepared in the presence of 0.5% Triton X-100. Protein size markers (K) shown on the left.

METHODS. *a*, Plasmids pTM-CFTR-3 and pTM-CFTR-3ΔF508 are derivatives of pSC-CFTR-2 and pSC-CFTR-2ΔF508 (Fig. 1) in which the T7 RNA promoter and EMC virus internal ribosome entry sequence from plasmid pTM-1 were placed immediately upstream of the CFTR initiation codon and the T7 terminator from pTM-1 was placed downstream of CFTR nucleotide 5,576 (refs 13, 17). For *in vitro* synthesis of full-length CFTR (1,480 amino acids), plasmids were linearized at a *Sall* site located downstream of the CFTR sequences and transcribed with T7 RNA polymerase (Stratagene). A truncated form of CFTR was synthesized from plasmids linearized with *Tth*1111, which restricts within codon 1,337 of the CFTR sequence. One μ g of each



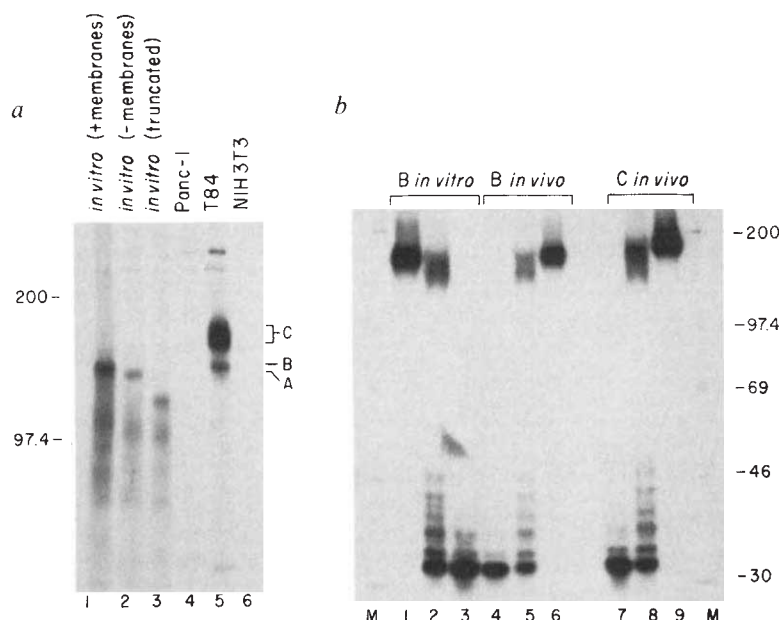
transcript was translated in a nuclease-treated rabbit reticulocyte lysate system (Promega) in a volume of 50 μl . Where indicated, 0.5% Triton X-100 or 7.2 equivalents of canine pancreatic microsomal membranes (Promega) were included in the reaction. *N*-Glycanase treatment of *in vitro* translation products was performed on 2 μl of translate (Genzyme). Reactions were diluted with an equal volume of sample buffer and 1 μl was electrophoresed, except in the case of the *N*-Glycanase reaction, in which 15 μl was analysed. Unless indicated, samples were left at room temperature for 10 min before electrophoresis and fluorography. The gel was exposed to film for 16 h. *b*, HeLa cells were plated at 50% confluency in 35-mm dishes 24 h before infection with recombinant vaccinia virus vT7-3 (ref. 13) (multiplicity of infection 10–20) and transfected with plasmids (5 μg) using 20 μg lipofectin³⁰ (BRL). Thirteen hours post-transfection, cells were labelled with [³⁵S]methionine (25 $\mu\text{Ci ml}^{-1}$) for 1 h in hypertonic media (190 mM NaCl) before collecting in 250 μl 135 mM NaCl, 20 mM Tris-HCl (pH 7.4), 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 100 $\mu\text{g l}^{-1}$ aprotinin and 17 μM PMSF. HeLa lysate (20 μl) or 1 μl of *in vitro* translate were mixed with an equal volume of sample buffer and analysed as above. To produce equivalent exposures, lanes 1–4 were exposed for 48 h and lanes 5 and 6 for 6 h.

FIG. 3 Immunoprecipitation of CFTR synthesized *in vitro* in rabbit reticulocyte lysates and *in vivo* in HeLa cells. A Plasmid pTM-CFTR-3 was transcribed with T7 RNA polymerase and translated in a rabbit reticulocyte lysate in the presence of canine microsomal membranes and [³⁵S]methionine. Samples were immunoprecipitated with non-immune (NI) rabbit serum (lane 1), a control irrelevant polyclonal antibody (CFN-21) raised against the *c-fyn* gene product¹⁹ (lane 2), or polyclonal antibody Ex13 (lanes 3, 4, 5 and 6). Similarly, immunoprecipitates derived from antibody Ex13 were prepared from lysates of [³⁵S]methionine-labelled HeLa cells infected with vaccinia virus and transfected with either pTM-1 (lane 9) or pTM-CFTR (lanes 7 and 8). Some reactions contained 5 μg (lane 4) or 10 μg (lanes 5 and 8) β-galactosidase-exon 13 fusion protein or 10 μg β-galactosidase-exon 10 fusion protein (lane 6). Fluorography was for 12 h. *b*, [³⁵S]methionine labelled CFTR synthesized *in vitro* and immunoprecipitated with antibody Ex13 (lane 1), Ex10 (lanes 2–4), monoclonal antibodies 13–1 (lanes 5–7) and 13–2 (lanes 8–10) raised against the β-galactosidase-exon 1 irreversible monoclonal antibody (423) raised to SV. Lanes 3, 7 and 10 contained exon 10 fusion pro-



6 and 9, exon 13 fusion protein (10 μ g). Fluorography was for 12 h. Preparation and purification of antibodies, and of fusion proteins were as previously published^{18,19,31–33}. Protein size markers (K) shown on the left.

FIG. 4 *a*, Immunoprecipitation of CFTR from nonrecombinant cells. Full-length CFTR synthesized *in vitro* in the presence (lane 1) or absence (lane 2) of membranes and truncated CFTR synthesized *in vitro*, terminating at residue 1,336 (lane 3) were immunoprecipitated with antibody Ex13 and incubated with protein kinase A (20 ng) and [γ - 32 P]ATP (10 μ Ci) in a final volume of 50 μ l in kinase buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂ and 100 μ g ml⁻¹ bovine serum albumin) at 30 °C for 60 min²². Immunoprecipitates of equivalent amounts of extracts of Panc-1 (lane 4), T84 (lane 5) and NIH 3T3 fibroblasts (lane 6) were subjected to the same treatment. Samples were heated at 37 °C for 10 min before electrophoresis. Exposure time was 20 min at -70 °C. *b*, One-dimensional peptide analysis of CFTR synthesized *in vitro* and *in vivo*. The CFTR synthesized in reticulocyte lysate in the presence of membranes (band B) and labelled with protein kinase A and [γ - 32 P]ATP (lanes 1-3) and that immunoprecipitated from T84 lysates (lanes 4-6, Band B and lanes 7-9, Band C) were digested with increasing amounts of *S. aureus* V8 protease¹⁹. Proteins in lanes 2, 5 and 8 were digested with 0.017 μ g μ l⁻¹ of *S. aureus* V8 protease, lanes 3, 4 and 7 with 0.17 μ g μ l⁻¹ of enzyme. Lanes 1, 6 and 9 were untreated samples. Exposure time was 24 h.



synthesized in HeLa cells transfected with CFTR cDNA. Bands A and B are efficiently immunoprecipitated by polyclonal antibody Ex13 (lanes 3 and 7). Neither preimmune serum nor immunoaffinity purified antibody raised against an irrelevant fusion protein (β -galactosidase-c-fyn; ref. 19) recognize the protein (lanes 1 and 2). Furthermore, the interaction is reversed by addition of excess β -galactosidase-exon 13 fusion protein (lanes 4, 5, 8), but not by an excess of irrelevant fusion protein (lane 6). These data show that the antibody specifically recognizes both the proteins synthesized *in vitro* and the recombinant cell-derived proteins.

Figure 3b shows immunoprecipitation of CFTR using two monoclonals (13-1 and 13-2) and immunoaffinity-purified polyclonal antibody Ex10. Band B is recognized efficiently and specifically. In all cases, addition of the appropriate fusion protein inhibits immunoprecipitation. An irrelevant control monoclonal antibody (raised against SV40 T-antigen; ref. 20) does not recognize the protein. Epitope mapping of the two monoclonals²¹ indicates that they recognize non-overlapping regions within exon 13 (not shown). Because two polyclonal antibodies and two distinct monoclonals raised against predicted CFTR sequences recognize the protein made *in vitro* and *in vivo*, there can be little doubt that the 130/135 K protein represents the product of the CFTR cDNA.

Our initial attempts to immunoprecipitate CFTR protein from metabolically labelled nonrecombinant cells were unsuccessful, probably because the amount of protein in these cells is small. We attempted to overcome the problem by phosphorylating the immunoprecipitated protein using the catalytic subunit of protein kinase A and [γ - 32 P]ATP (ref. 22). Figure 4 shows proteins synthesized *in vitro* and labelled with 32 P after immunoprecipitation with the polyclonal antibody Ex13. Lane 1 shows that a protein migrating in the position of band B is labelled in immunoprecipitates of material synthesized *in vitro* in the presence of canine microsomal membranes and full-length CFTR cDNA. More rapidly migrating proteins are detected in immunoprecipitates of the cell-free product made in the absence of membranes (lane 2) or under the direction of the truncated cDNA (lane 3). These results indicate that CFTR synthesized *in vitro* is a good substrate for protein kinase A when it is present in immunoprecipitates.

Immunoprecipitates of extracts of nonrecombinant cells were labelled in the same way. Lane 5 shows that 32 P-labelled material comigrating with band B was detected in immunoprecipitates of T84 cells²³ treated with protein kinase A. A more highly

phosphorylated, diffuse band of $M_r \sim 150$ K (band C) was also present. T84 cells are a human colon carcinoma cell line known to contain relatively large amounts of CFTR messenger RNA⁷. Bands comigrating with B and C were not detected in Panc-1 cells²⁴ (lane 4), mouse 3T3 cells (lane 6) or 27Sk human fibroblasts (not shown). It is believed that these cells do not express CFTR⁷. Bands comigrating with bands B and C were also detected in similar experiments using extracts of T84 cells that had been reacted with the antibodies Ex10, 13-1 and 13-2 (not shown).

The distribution in different cell types and the finding that four antibodies against CFTR sequences recognize material of similar mobility, suggests that the bands B and C labelled in T84 cell immunoprecipitates are CFTR. To confirm this, we excised 32 P-labelled band B from immunoprecipitates of material synthesized in the cell-free system and prepared a partial proteolysis fingerprint using *Staphylococcus aureus* V8 protease. In parallel, we prepared similar fingerprints of bands B and C from 32 P-labelled immunoprecipitates of T84 cells. As seen in Fig. 4b, the partial proteolysis fingerprints are virtually identical, demonstrating that the proteins immunoprecipitated from T84 cells are related to CFTR synthesized *in vitro*. Similar conclusions were drawn from experiments using endoproteases Asp-N and Lys-C (not shown). One interpretation of this result is that band C represents a fully processed version of the CFTR protein, with the addition of more carbohydrate or phosphate, or both, to the core-glycosylated CFTR. The related multiple drug resistance P-glycoprotein undergoes a similar shift in mobility on addition of N-linked carbohydrate¹⁶.

These studies constitute a formal identification of the product of the CFTR open reading frame and verify several of the predictions made about the CFTR protein. First, CFTR associates with membranes and is glycosylated, properties consistent with the finding that the CF defect in Cl⁻ channel regulation is observed in isolated cell-free patches of membrane. Second, CFTR can be phosphorylated by cAMP-dependent protein kinase, consistent with the finding that cAMP regulates Cl⁻ channels in normal but not CF cell membranes. Third, CFTR is present in T84 cells but not fibroblasts, consistent with the finding that T84 cells show Cl⁻ channel regulation similar to that observed in normal airway epithelial cells, whereas channel regulation in fibroblasts is controversial. Furthermore, the agents described here provide materials for a wide range of studies on the CFTR and its function: for example, in the accompanying

paper²⁵ we describe complementation of the defect in ion transport in CF epithelial cells by transfection of CFTR cDNA. Of particular interest will be the mechanism by which the biochemical activity of CFTR regulates, either directly or indirectly, the absorption and secretion of chloride ion. In terms of potential therapy for CF, the system we describe offers a means to screen for pharmacological compounds that interact with CFTR, and to test protein replacement and gene therapy as approaches in the treatment of CF. □

Received 27 July; accepted 31st August 1990.

1. Boat, T., Welsh, M. & Beaudet, A. in *The Metabolic Basis of Inherited Disease* (eds Scriver C., Beaudet A., Sly, W. & Valle, D.) 2649–2680 (McGraw Hill, New York, 1989).
2. Quinton, P. *Clin. Chem.* **35**, 726–730 (1989).
3. Li, M. *et al. Nature* **331**, 358–360 (1988).
4. Frizzell, R., Reckemmer, G. & Shoemaker, R. *Science* **233**, 558–560 (1986).
5. Welsh, M. *Science* **232**, 1648–1650 (1986).
6. Rommens, J. *et al., Science* **245**, 1059–1065 (1989).
7. Riordan, J. *et al., Science* **245**, 1066–1073 (1989).
8. Kerem, B.-S. *et al. Science* **245**, 1073–1080 (1989).
9. Brosius, J. *Gene* **27**, 151–160 (1984).
10. Hawley, D. K. & McClure, W. R. *Nucleic Acids Res.* **11**, 2237–2255 (1983).
11. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3240–3244 (1973).
12. Jang, S. K., Davies, M. V., Kaufman, R. J. & Wimmer, E. *J. Virol.* **63**, 1651–1660 (1989).

13. Elroy-Stein, O., Fuerst, T. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6126–6130 (1989).
14. Greenberger, L. M., Williams, S. S. & Horowitz, S. B. *J. biol. Chem.* **262**, 13685–13689 (1987).
15. Nash, B. & Tate, S. S. *J. biol. Chem.* **259**, 678–684 (1984).
16. Richard, N. D., Aldwin, L., Nitecki, D., Gottesman, M. & Pastan, I. *Biochemistry* **27**, 7607–7613 (1988).
17. Fuerst, T. R., Niles, E. G., Studier, W. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8122–8126 (1986).
18. Kohler, G. & Milstein, C. *Nature* **256**, 495–497 (1975).
19. Cheng, S. H. *et al. EMBO J.* **7**, 3845–3855 (1988).
20. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **39**, 861–869 (1981).
21. Wurznner, R., Oppermann, M., Zierz, R., Baumgarten, H. & Gotze, O. *J. Immun. Meths.* **126**, 231–237 (1980).
22. Kawata, M. *et al. J. biol. Chem.* **264**, 15688–15695 (1989).
23. Muriakarmi, H. & Masui, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3464–3468 (1980).
24. Lieber, M., Mazetta, J., Nelson-Rees, W., Kaplan, M. & Todaro, G. *Int. J. Cancer* **15**, 741–747 (1975).
25. Rich, D. P. *et al. Nature* **347**, 358–363 (1990).
26. Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
27. Stoker, N. G., Fairweather, N. F. & Spratt, B. G. *Gene* **18**, 335–341 (1982).
28. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
29. Tarentino, A. L., Gomez, C. M. & Plummer, T. H. *Biochemistry* **24**, 4665–4671 (1985).
30. Feigner, P. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 7413–7417 (1987).
31. Harlow, E. & Lane, D. In *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1988).
32. Harvey, R., Hehir, K. M., Smith, A. E. & Cheng, S. H. *Molec. cell. Biol.* **9**, 3647–3656 (1989).
33. Mole, S. & Lane, D. *J. Virol.* **54**, 703–710 (1985).

ACKNOWLEDGEMENTS. We thank G. White for oligonucleotides; G. Horn and D. Dearborn (Cleveland) for advice and reagents in the early stages of this work; B. Moss (NIH) for the plasmid pTM-1; H. de Jonge (Rotterdam) for advice, and for exchanging antibodies; C. O'Riordan, D. Souza, J. McPherson and R. Douglas (Genzyme) for advice and help, and B. Herrera for secretarial assistance. This work was supported in part by the CF Foundation and the National Heart, Lung and Blood Institute.

Inhibition by dopamine of (Na⁺ + K⁺)ATPase activity in neostriatal neurons through D₁ and D₂ dopamine receptor synergism

Alejandro M. Bertorello*, Jessica F. Hopfield†, Anita Aperia* & Paul Greengard†‡

* Department of Pediatrics, St Göran's Children's Hospital, Karolinska Institutet, S-11281 Stockholm, Sweden

† Laboratory of Molecular and Cellular Neuroscience,

The Rockefeller University, New York, New York 10021, USA

THE (Na⁺ + K⁺)ATPase, an integral membrane protein located in virtually all animal cells, couples the hydrolysis of ATP to the countertransport of Na⁺ and K⁺ ions across the plasma membrane^{1,2}. In neurons, a large portion of cellular energy is expended by this enzyme to maintain the ionic gradients that underlie resting and action potentials. Although neurotransmitter regulation of the enzyme in brain has been reported^{3,4}, such regulation has been characterized either as a nonspecific phenomenon^{5,6} or as an indirect effect of neurotransmitter-induced changes in ionic gradients⁷. We report here that the neurotransmitter dopamine, through a synergistic effect on D₁ and D₂ receptors, inhibits the (Na⁺ + K⁺)ATPase activity of isolated striatal neurons. Our data provide unequivocal evidence for regulation by a neurotransmitter of a neuronal ion pump. They also demonstrate that synergism between D₁ and D₂ receptors, which underlies many of the electrophysical and behavioural effects of dopamine in the mammalian brain⁸, can occur on the same neuron. In addition, the results support the possibility that dopamine and other neurotransmitters can regulate neuronal excitability through the novel mechanism of pump inhibition.

The activity of (Na⁺ + K⁺)ATPase was assayed as ouabain-sensitive ATP hydrolysis in a cell suspension of acutely dissociated striatal neurons from guinea pig. Dissociated striatum yields a cell population which, morphologically, is nearly homogeneous: an estimated 95% of striatal neurons are of a single type, namely medium spiny type I (ref. 9), and almost

all astrocytes and myelinated fibres settle out of the cell suspension. Isolated cells retain their characteristic medium spiny morphology and proximal dendritic structure and display voltage- and neurotransmitter-coupled ionic conductances (J.F.H., unpublished observations). The (Na⁺ + K⁺)ATPase activity in control samples averaged 199 ± 13 nmol Pi per mg protein per min (*n* = 19). Ouabain-insensitive ATP hydrolysis (Mg²⁺ ATPase activity) averaged 209 ± 14 nmol Pi per mg protein per min (*n* = 19) and was unaffected by any of our experimental treatments.

Inhibition by dopamine of (Na⁺ + K⁺)ATPase activity is shown as a function of dopamine concentration and time in Fig. 1. Half-maximal inhibition was achieved at 10⁻⁹ M dopamine and maximal inhibition at 10⁻⁶ M. The inhibitory effect was maximal within 5 min of incubation and persisted for at least 15 min.

To analyse the contributions of D₁ and D₂ dopamine receptor subtypes in mediating this effect of dopamine, it was necessary to prevent activation of receptors by endogenous dopamine. Animals were therefore pretreated with the L-amino-acid dopa decarboxylase inhibitor, benserazide, to deplete endogenous dopamine stores¹⁰ (see Fig. 1 legend). In benserazide-treated animals, neither 100 nM of the D₁ selective agonist SKF 82526 nor 100 nM of the D₂ selective agonist LY 171555 was itself sufficient to produce significant inhibition of (Na⁺ + K⁺)ATPase activity (Fig. 2). Addition of 100 nM SKF 82526 with 100 nM LY 171555 resulted in an inhibition to ~50% of control level, suggesting that both receptor subtypes must be activated to induce enzyme inhibition. The synergistic effect of D₁ and D₂ receptor coactivation was not a reflection of higher total agonist concentration in the agonist mixture: 1 µM doses of SKF 82525 and LY 171555 resulted in (Na⁺ + K⁺)ATPase activity of 104 ± 38% (*n* = 4) and 90 ± 10% (*n* = 2) of control values, respectively.

In vehicle-treated animals that did not receive benserazide (*n* = 3), 100 nM SKF 82526 and 100 nM LY 171555 each given alone reduced (Na⁺ + K⁺)ATPase activity to 78 ± 7% and 70 ± 13% of control levels, respectively. The two agonists, when present together, reduced enzyme activity to 65 ± 8% of control. This ability of either selective agonist alone to inhibit (Na⁺ + K⁺)ATPase activity suggests that endogenous dopamine in the cell suspension, released during mechanical trituration of the slices, may have caused some activation of both receptor subtypes. In support of this interpretation, basal enzyme activity in samples from vehicle-treated animals was slightly lower (154 ± 22 nmol Pi per mg protein per min, *n* = 3) than in samples from benserazide-treated animals (179 ± 10 nmol Pi per mg protein

‡ To whom correspondence should be addressed.

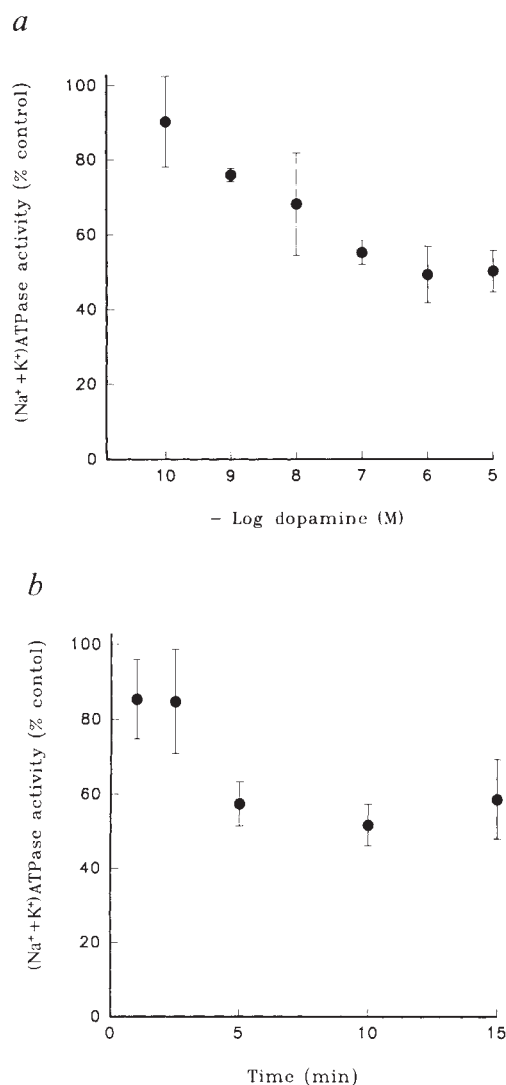


FIG. 1 Inhibition of (Na⁺ + K⁺)ATPase activity by dopamine. **a**, Cells were incubated for 10 min with the indicated concentrations of dopamine. **b**, Cells were incubated with 1 μ M dopamine for the indicated periods of time. All samples were compared with control samples incubated without dopamine for 10 min (activity = 206 ± 15 nmol Pi per mg protein per min, $n = 28$). Data represent means $\pm$ s.e.m. of three to seven experiments. No change in enzyme activity was observed in samples of control cells incubated without dopamine for up to 15 min.

METHODS. Acutely dissociated striatal neurons from guinea pig brain were prepared by trypsin digestion as described²⁵, with the following modifications: coronal slices from left and right striata of 120–175 g guinea pigs were incubated with 1 mg ml⁻¹ trypsin at 30 °C for 1.25 h. After a 30 min incubation at 22 °C in fresh PIPES buffered saline, slices were triturated in 1 ml of Ca²⁺-free, 2 mM Mg²⁺, PIPES saline. (Na⁺ + K⁺)–ATPase activity was measured as previously described¹³ with the following modifications: preincubation was started by transferring 45- μ l aliquots of the cell suspension to HEPES buffered saline containing the dopaminergic drugs or vehicle (final volume 100 μ l). The preincubation period was stopped by rapid cooling to 4 °C. Aliquots (10 μ l) of the cell suspension (protein concentration 2–7 μ g protein) were transferred to the (Na⁺ + K⁺)ATPase assay medium (final volume 100 μ l) containing, NaCl 50 mM, KCl 5 mM, MgCl₂ 10 mM, EGTA 1 mM, Tris-HCl 100 mM and [³²P]Na₂ATP 10 mM (specific activity 3,000 Ci mmol⁻¹). Cells were transiently permeabilized by placing the samples at –20 °C for 10 min to permit the entry of [³²P]ATP and stabilize the intracellular Na⁺ concentration at 70 mM (50 mM NaCl and 10 mM [³²P]Na₂ATP). Cells did not freeze during this procedure and they maintained their morphology. Samples were then incubated at 37 °C for 15 min. The reaction was stopped by rapid cooling to 4 °C and TCA/charcoal (5%/10%) was added immediately. (Na⁺ + K⁺)ATPase activity, measured by counting the liberated ³²P, was calculated as the difference between the averages of duplicate test samples (total ATPase activity) and duplicate samples assayed in a Na⁺- and K⁺-free solution containing 1 mM ouabain (ouabain-insensitive ATPase activity). Protein was measured by the method of Bradford²⁶. For the experiments involving dopamine-depleted animals, guinea pigs were anaesthetized with methoxyflurane and tube fed 300 mg benserazide chloride per kg bodyweight using 1 ml water as vehicle, 1 h before being killed. For measurement of dopamine levels, tissue samples were homogenized in 0.4 M perchloric acid (PCA), and the dopamine adsorbed onto alumina at pH 8.6 and eluted in 0.1 M PCA. Dopamine in the eluate was analysed by reverse-phase HPLC with electrochemical detection²⁷ using a BAS LC-4B detector and a BAS PM-48 pump. Protein was analysed by the procedure of Lowry *et al.*²⁸. Trypsin digested striatal slices from benserazide-treated animals had a mean dopamine concentration of 101 ± 16 pmol dopamine per mg protein ($n = 4$) compared with 235 ± 37 pmol dopamine per mg protein ($n = 3$) for vehicle-treated animals. Data are presented as means $\pm$ s.e.m. and were analysed by Student *t*-tests.

per min, $n = 4$), possibly reflecting tonic inhibition of enzyme activity by endogenous dopamine in the cell suspensions.

To investigate further the pharmacology of this inhibition, cells from animals not treated with benserazide were incubated with dopamine, in the absence or presence either of the selective D₁ antagonist SCH 23390 or of one of the selective D₂ antagonists, YM-09151 or S-sulpiride (Fig. 3). Addition of either a D₁ or D₂ specific antagonist (10 μ M) to the cell suspension was sufficient to prevent inhibition induced by 1 μ M dopamine of (Na⁺ + K⁺)ATPase activity.

Dopamine regulates (Na⁺ + K⁺)ATPase activity in permeabilized rat renal cells, but has no effect on the activity of the purified enzyme from rat kidney^{11–15}. It has been difficult to interpret experiments reporting regulation of enzyme activity by neurotransmitters in brain for two reasons. First, in broken cell preparations, catecholamines seem to enhance (Na⁺ + K⁺)ATPase activity nonspecifically by chelating divalent metal ions that normally inhibit the enzyme⁵. Second, activity of the enzyme is dependent on and extremely sensitive to the Na⁺ and K⁺ concentrations near the cell membrane, which might be altered as a result of neurotransmitter binding and ion channel activation in intact cells (for example, ref. 16). In the present experiments, intact cells were preincubated with dopamine agonists or antagonists, then permeabilized by cold shock to load them with [³²P]ATP and stabilize intracellular Na⁺ concentration at 70 mM. This procedure, which did not disrupt cell morphology, should have dissipated any neurotransmitter-

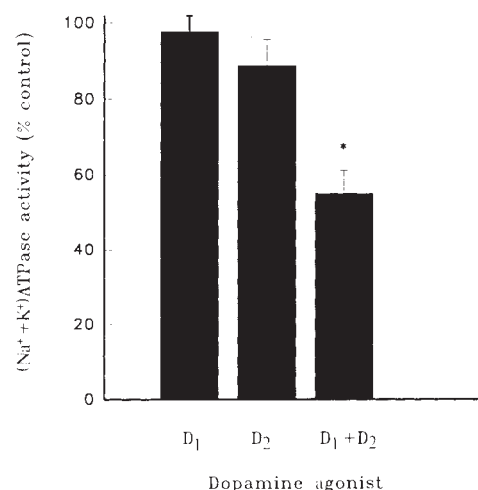


FIG. 2 Synergistic action of D₁ and D₂ agonists as inhibitors of (Na⁺ + K⁺)ATPase activity. All animals had been pretreated with benserazide to deplete endogenous dopamine. D₁, 100 nM of the D₁ agonist SKF 82526; D₂, 100 nM of the D₂ agonist LY 171555. Results represent means $\pm$ s.e.m. of four experiments. In each experiment, results were calculated as percentage of a non-agonist-treated control. *, $P < 0.005$ compared with control and compared with D₁ and D₂ agonists alone.

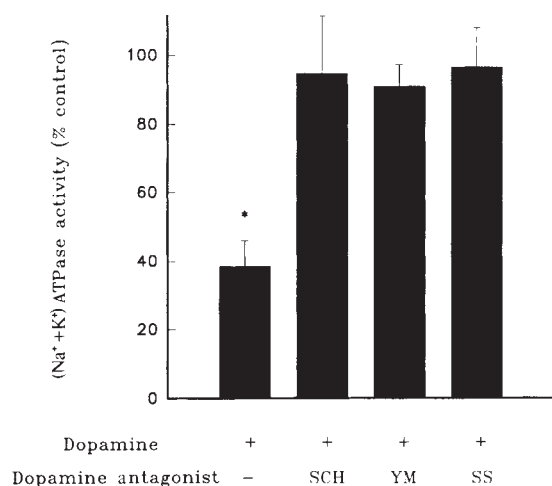


FIG. 3 D₁ and D₂ antagonists each prevent the inhibitory action of 1 μ M dopamine. SCH, 10 μ M of the D₁ antagonist SCH 23390; YM, 10 μ M of the D₂ antagonist YM-09151; SS, 10 μ M of the D₂ antagonist S-sulpiride. Results represent means $\pm$ s.e.m. of four experiments. For each antagonist, data represent enzyme activity in the presence of dopamine expressed as percentage of that observed in the absence of dopamine (mean activity = 147 ± 11 nmol Pi per mg protein min⁻¹, $n=14$). *, $P < 0.005$ compared with control (no dopamine) and compared with dopamine plus dopamine antagonists.

induced changes in ion concentrations which might have occurred before enzyme activity was measured.

An understanding of the mechanism of action of dopamine in the brain is important in view of the widespread belief that this neurotransmitter plays an important part in the aetiology of several neuropsychiatric disorders including schizophrenia¹⁷. The micromolar concentrations of D₁ and D₂ agonists used here to inhibit (Na⁺ + K⁺)ATPase activity indicate that the observed regulation of enzyme activity by dopamine may have physiological significance. Inhibition of the electrogenic (Na⁺ + K⁺)ATPase leads to transient membrane depolarization and a rise in intracellular Na⁺ in excitable cells³. Therefore, the present results suggest that dopamine, in addition to its known regulation of ion channels¹⁸, regulates cell excitability through the novel mechanism of pump inhibition. If this is so, then this action of dopamine on (Na⁺ + K⁺)ATPase activity may be an important component of the mechanism by which this neurotransmitter alters the responsivity of dopaminergic neurons to other neurotransmitters¹⁹.

The present results support the idea that activation of both D₁ and D₂ receptors is required for inhibition of (Na⁺ + K⁺)ATPase activity. D₁ and D₂ receptor subtypes in brain differ in their anatomical distributions²⁰ and pharmacological profiles²¹, and mediate distinct behavioural and electrophysiological effects²². An interaction between these receptor subtypes has been extensively described. Synergism has previously been demonstrated on both the behavioural and electrophysiological levels (for review see ref. 8). The coactivation of both receptor subtypes required for many postsynaptic dopamine-mediated effects could reflect interaction of receptors on a single cell, on different cells from the same brain region or on cells from different brain regions. The present study used a dissociated cell population, nearly homogeneous morphologically, and lacking synaptic contacts. D₁ and D₂ receptor coactivation resulted in a synergistic inhibition of (Na⁺ + K⁺)ATPase activity distinct from the additive effect expected from subpopulations of neurons each containing only one of the two receptor subtypes. Therefore, the synergism demonstrated here must have occurred at the level of single cells.

The biochemical mechanisms underlying inhibition of (Na⁺ + K⁺)ATPase activity by dopamine represent an important area

for future research. D₁ receptor activation in striatum stimulates adenylate cyclase activity²³ leading to the formation of cyclic AMP and the activation of cAMP-dependent protein kinase²⁴. Preliminary experiments with the potent adenylate cyclase activator forskolin support the involvement of this signal transduction pathway in the D₁ receptor contribution to inhibition of (Na⁺ + K⁺)ATPase activity: coapplication of the D₂ agonist LY 171555 with forskolin caused a synergistic reduction of (Na⁺ + K⁺)ATPase activity (unpublished observations). It will be interesting to determine the signal transduction pathway used by the D₂ receptor in contributing to the inhibition of (Na⁺ + K⁺)ATPase activity as well as the molecular basis for the synergism between the D₁ and D₂ pathways observed here. □

Received 8 May; accepted 9 July 1990.

- Skou, J. C. *Physiol. Rev.* **45**, 596–617 (1965).
- Glynn, I. M. in *The Enzymes of Biological Membranes* (ed. Martonosi, A. N.) 35–114 (Plenum, New York, 1985).
- Phillips, J. W. & Wu, P. H. *Prog. Neurobiol.* **17**, 141–184 (1981).
- Smith, P. A. *Trends Pharmacol. Sci.* **422**–425 (1984).
- Van der Krogt, J. & Belfroid, R. D. M. *Biochem. Pharmacol.* **29**, 857–868 (1980).
- Lafferty, P., Jackson, D. M. & Malor, R. *Biochem. Pharmacol.* **34**, 3591–3596 (1985).
- Smith, P. A., Thompson, E. L. & Zidichouski, J. A. *Neurosci. Lett.* **71**, 72–76 (1986).
- Clark, D. & White, F. J. *Synapse* **1**, 347–388 (1987).
- Gerfen, C. J. *Electron Microsc. Tech.* **10**, 265–281 (1988).
- Da Prada, M., Keller, H. H., Pieri, L., Kettler, R. & Haefely, W. E. *Experientia* **40**, 1165–1171 (1984).
- Aperia, A., Bertorello, A. & Seri, I. *Am. J. Physiol.* **252**, F39–F45 (1987).
- Bertorello, A. & Aperia, A. *Acta physiol. scand.* **132**, 441–443 (1988).
- Bertorello, A. & Aperia, A. *Am. J. Physiol.* **256**, F57–F69 (1989).
- Bertorello, A. & Aperia, A. *Am. J. Physiol.* **256**, F370–F373 (1989).
- Bertorello, A. & Aperia, A. *Acta physiol. scand.* **130**, 571–574 (1987).
- Gadsby, D. C. *Nature* **306**, 691–693 (1983).
- Carlsson, A. *Neuropsychopharmacology* **1**, 179–186 (1988).
- Freedman, J. E. & Weight, F. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3618–3622 (1988).
- Mercuri, N. et al. *Brain Res.* **385**, 110–121 (1985).
- Richfield, E. K., Penny, J. B. & Young, A. B. *Neuroscience* **30**, 767–777 (1989).
- Stoof, J. C. & Kebabian, J. W. *Life Sci.* **35**, 2281–2296 (1984).
- Wang, R. Y., White, F. J., Mereu, G. & Hu, X. T. in *Dopamine Receptors* (ed. Creese, I.) 153–173 (Alan Liss, New York, 1987).
- Kebabian, J. W., Petzold, G. L. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2145–2149 (1972).
- Hemmings, H. C., Walaas, I. S., Ouimet, C. C. & Greengard, P. in *Receptor Biochemistry and Methodology, Vol. 9: Structure and Function of Dopamine Receptors* (eds Creese, I. & Fraser, C. M.) 115–151 (Liss, New York, 1987).
- Kay, A. R. & Wong, R. K. S. *J. Neurosci. Meth.* **16**, 227–238 (1986).
- Bradford, M. M. *Analyt. Biochem.* **72**, 248–254 (1976).
- Sharp, T., Zetterstrom, T. & Ungerstedt, U. *J. Neurochem.* **47**, 113–122 (1986).
- Lowry, O. K., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. Biol. Chem.* **193**, 265–275 (1951).

ACKNOWLEDGEMENTS. We thank G. Fried for performing the HPLC dopamine measurements and D. Gadsby for advice and critical comments. SKF 82526 was a gift from Smith, Kline & French. This work was supported by the Swedish Medical Research Council (A.B. and A.A.) and the USPHS and the USEPA (J.F.H. and P.G.).

Activation of latent Ca²⁺ channels in renal epithelial cells by parathyroid hormone

Brian J. Bacskai*† & Peter A. Friedman‡

Thayer School of Engineering, Dartmouth College*, and Department of Pharmacology and Toxicology, Dartmouth Medical School, Hanover, New Hampshire 03756, USA

CALCIUM has an important role in regulating epithelial cell ion transport^{1,2} and is itself transported by tissues involved in the maintenance of extracellular Ca²⁺ homeostasis^{3–6}. Although the mechanism of Ca²⁺ entry in electrically excitable cells is well-documented^{7,8} little is known about it in epithelial cells. Calcium absorption in polarized epithelial cells is a two-step process in which Ca²⁺ enters cells across apical plasma membranes and is extruded across basolateral membranes⁹. Efflux may be mediated by an energy-dependent Ca²⁺-ATPase or by Na⁺/Ca²⁺ exchange^{10–12}. We examined Ca²⁺ influx in single cultured cells

* Present address: Howard Hughes Medical Institute, University of California–San Diego, La Jolla, California, USA.

† To whom correspondence should be addressed.

from distal renal tubules sensitive to parathyroid hormone by measuring intracellular Ca^{2+} . Our results demonstrate that parathyroid hormone activates dihydropyridine-sensitive channels responsible for Ca^{2+} entry. We also show that microtubule-dependent exocytosis stimulated by parathyroid hormone may be necessary for the insertion or activation of Ca^{2+} channels in these cells. Once inserted or activated, dihydropyridine-sensitive channels mediate Ca^{2+} entry into these Ca^{2+} -transporting epithelial cells. Our results support the view that agonist-induced exocytosis may represent a general paradigm for modulation of transport in epithelial cells by delivery and incorporation of transport proteins to plasma membranes or by delivery to plasma membranes of factors regulating these proteins.

The effect of parathyroid hormone (PTH) on intracellular Ca^{2+} activity ($[\text{Ca}^{2+}]_i$) in a single cell is depicted in Fig. 1. Before PTH addition, intracellular $[\text{Ca}^{2+}]_i$ was ~ 100 nM and was nearly uniform throughout the cytoplasm. After 30 min of exposure to the 1–34 amino-terminal fragment of bovine PTH (bPTH(1–34)), $[\text{Ca}^{2+}]_i$ increased by more than 200%. A time-course of the changes of $[\text{Ca}^{2+}]_i$ in a single cell is shown in Fig. 2a. Addition of PTH led to a slow increase in $[\text{Ca}^{2+}]_i$ (10–12-min latency, 30-min maximum response) from resting levels of

100 nM to sustained values of about 250 nM. A comparable time-course and response to PTH was also observed in a mouse cortical ascending limb (CAL) microperfused *in vitro* and maintained at 37 °C (Fig. 2b). The similarity of time-courses in an intact tubule and in the cultured cells indicates that the latency of the increase in $[\text{Ca}^{2+}]_i$ after addition of PTH represents a physiological delay, and does not result from diffusion constraints presented by this cell culture preparation, temperature effects, or from inadequate mixing. The time-course of the PTH-stimulated rise in $[\text{Ca}^{2+}]_i$ in CAL+distal convoluted tubule (DCT) cells, however, is qualitatively unlike that in renal proximal tubule cells^{13,14} or in cultured osteoblasts¹⁵. In these latter cell types, PTH elicits a rapid but transient increase in $[\text{Ca}^{2+}]_i$. Such $[\text{Ca}^{2+}]_i$ transients are generally associated with signal transduction resulting from mobilization of Ca^{2+} from intracellular stores^{16,17}.

After the PTH-induced increase in $[\text{Ca}^{2+}]_i$ reached plateau levels, the Ca^{2+} channel agonist Bay k 8644 was added to the incubation medium to determine if dihydropyridine-sensitive Ca^{2+} channels mediate Ca^{2+} entry in these cells. Bay k 8644 elicited an immediate further rise in $[\text{Ca}^{2+}]_i$ to about 450 nM (Fig. 3a). But Bay k 8644 treatment alone had no effect on

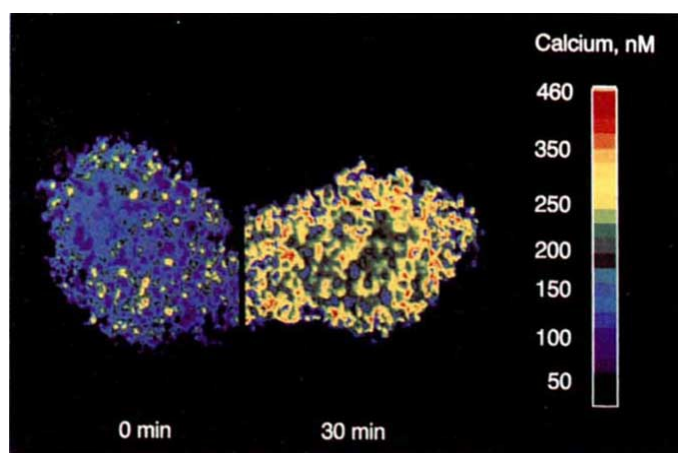
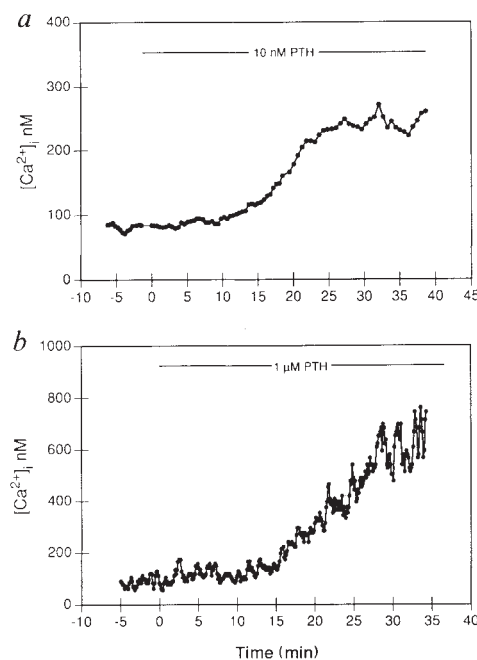


FIG. 1 PTH effect on $[\text{Ca}^{2+}]_i$ in a single CAL+DCT cell. PTH-sensitive CAL+DCT cells were isolated from mouse kidneys by a double antibody method. After mechanical and chemical (0.1% collagenase) digestion of kidney cortex to single cells, the resulting tissue was incubated with goat anti-human Tamm-Horsfall glycoprotein (THGP) antisera (Cappel, Malvern, Pennsylvania)²⁹. The resulting material was incubated with coated ferrous particle-conjugated rabbit anti-goat IgG (Advanced Magnetics, Cambridge, Massachusetts). THGP-positive cells were separated from the mixture with a magnet. The final cell isolate was grown 4–7 days in culture on glass coverslips (31 mm, #00; Biophysica, Baltimore) in Opti-MEM medium (Gibco) at 37 °C in a humidified incubator equilibrated with 95% air/5% CO_2 . Intracellular Ca^{2+} activity ($[\text{Ca}^{2+}]_i$) in single cells was measured with fura-2 using a dual-excitation, digital imaging fluorescence microscope. Cells were incubated with fura-2-AM (Calbiochem, San Diego, California or Molecular Probes, Eugene, Oregon) (10 μM) for 45–60 min at 37 °C, rinsed, then mounted in a perfusion chamber (Biophysica) on an inverted microscope (Nikon Diaphot). The external bath consisted of a HEPES-buffered, balanced saline solution (CaCl_2 1.0 mM; NaCl 140 mM; HEPES 3 mM; MgCl_2 1.2 mM; urea 10 mM; PO_4 2.8 mM; KCl 5.1 mM), pH 7.4, 23–25 °C. A device allowing two wavelengths (340 and 380 nm) to be alternately selected using micro-computer-driven shutters permitted dual-excitation wavelength measurements. Cells were viewed with a $\times 40$ oil immersion fluorescence objective (Nikon). Fluorescence emission was monitored at 510 nm with an image intensifier (Opelco, Washington DC) coupled to a Newvicon-type video camera (Ikegami, ITC510, Opelco). Video output was digitized at 512×512 pixels by 256 gray levels per pixel with a video image processor (DR2851, Data Translation, Marlboro, Massachusetts) in an IBM PC/AT. The system was optimized for minimum excitation intensities and maximum emission collection efficiency to prevent photobleaching of fura-2. Each image was produced by dividing the average of 10 video frames at 340 nm by the average of 10 video frames at 380 nm on a pixel-by-pixel basis in a single cell before

FIG. 2 Time-course of PTH-stimulated increases of $[\text{Ca}^{2+}]_i$. Each point represents a determination of $[\text{Ca}^{2+}]_i$ within a region of interest within a single cell and was acquired as described in Fig. 1. a, PTH effect in a single cultured cell. The bPTH(1–34) (10 nM) was added at $t=0$ min. In five experiments, with each cell serving as its own control, PTH increased $[\text{Ca}^{2+}]_i$ from 105 ± 9 to 260 ± 14 nM (mean $\pm$ s.e.m., $P < 0.001$; paired t -test). b, PTH effect in a CAL nephron segment. To compare the results in cultured cells with those in intact, freshly isolated tissue, a mouse CAL was isolated and microperfused *in vitro* according to established methods^{32–34}. Fura-2 was loaded by adding fura-2-AM (10 μM) to the luminal perfusate for 20 min. The bPTH(1–34) (1 μM) was added at $t=0$ min.



and after a 30 min incubation with 10 nM bPTH(1–34) (Sigma, St. Louis). The concentration of PTH was selected on the basis of the K_m for activation of adenylyl cyclase by PTH(1–34) in these cells³⁰. The corresponding ratio values were scaled by a factor of 100 and the image was pseudo-coloured according to varying levels of $[\text{Ca}^{2+}]_i$, as indicated in the scale bar. Absolute values of $[\text{Ca}^{2+}]_i$ were based on calibrations performed after every experiment. Calibration was accomplished by adding 5 μM ionomycin to obtain the maximum fluorescence ratio, followed by addition of 10 mM EGTA to obtain the minimum fluorescence ratio. A dissociation constant K_d of fura-2 for Ca^{2+} of 224 nM was assumed³¹.

$[Ca^{2+}]_i$ (Fig. 3b). Subsequent addition of PTH after Bay k 8644 treatment increased $[Ca^{2+}]_i$ to levels equivalent to those in the former experiment. Again, a latency of 10–12 min preceded discernible increases in $[Ca^{2+}]_i$. These results are consistent with the view that in these cells PTH may lead to the recruitment of new dihydropyridine-sensitive Ca^{2+} channels or, alternatively, to the activation or unmasking of dihydropyridine binding sites of existing plasma membrane Ca^{2+} channels. The dihydropyridine Ca^{2+} -channel blocker nifedipine abolished the PTH stimulation of $[Ca^{2+}]_i$ but had no effect on its own (Fig. 3c). Moreover, the inhibitory effect of nifedipine was reversed after its replacement with equimolar concentrations of Bay k 8644 in the presence of PTH. The increase in $[Ca^{2+}]_i$ after removal of nifedipine and addition of Bay k 8644 was immediate, suggesting that the cascade of events leading to activation of Ca^{2+} entry

was initiated even when Ca^{2+} entry was prevented. Finally, the observation that nifedipine blocked PTH-induced increases of $[Ca^{2+}]_i$ supports the view that extracellular fluid represents the source of elevated $[Ca^{2+}]_i$. Further evidence for this comes from experiments in which Ca^{2+} was omitted from the extracellular bath. Under these conditions, PTH failed to elevate $[Ca^{2+}]_i$. After repletion of extracellular calcium, $[Ca^{2+}]_i$ increased promptly, demonstrating as before that in the absence of extracellular Ca^{2+} , the events initiated by PTH were activated, but no Ca^{2+} entry into cytoplasm occurred. These results also contrast with those for renal proximal tubule cells or cultured osteoblasts^{13–15}, in which the rapid increases of $[Ca^{2+}]_i$ after addition of PTH has its origin in Ca^{2+} release from intracellular stores.

In electrically excitable cells, depolarization of voltage-gated, dihydropyridine-sensitive Ca^{2+} channels stimulates Ca^{2+} entry¹⁸. In the absence of PTH, exposure of CAL+DCT cells to 50 mM K^+ —a treatment which depolarizes similar epithelial cells¹⁹, even in the presence of the K^+ ionophore valinomycin—had no detectable effect on $[Ca^{2+}]_i$. Subsequent addition of Bay k 8644 did not increase $[Ca^{2+}]_i$ (not shown). Although preliminary whole-cell patch experiments failed to indicate voltage sensitivity of PTH-activated Ca^{2+} channels in these cells (P. Dietl, B. A. Stanton and P.A.F., unpublished observations), complete characterization will be required to determine if these are conventional L-type Ca^{2+} channels or if they represent a tissue-specific channel isoform²⁰ unique to epithelia or to CAL+DCT cells in particular. The apparent absence of voltage sensitivity is not surprising in view of the fact that these polarized epithelial cells are not electrically excitable and may indicate that these dihydropyridine-sensitive Ca^{2+} channels are regulated by phosphorylation. The observed latency of the $[Ca^{2+}]_i$ response to PTH, and the absence of an increase in $[Ca^{2+}]_i$ using Bay k 8644 alone, suggests that few active Ca^{2+} channels are present in plasma membranes of these cells in the absence of PTH. Collectively, these observations suggest that the PTH-stimulated rise in $[Ca^{2+}]_i$ resulted from Ca^{2+} entry through dihydropyridine-sensitive channels and is consistent with the view that PTH leads to the recruitment of new, or activation of extant, plasma membrane Ca^{2+} channels.

Emerging evidence indicates that proteins such as water channels^{21,22}, proton pumps^{23,24} and chloride channels (R. Frizzel and D. Stetson, personal communications), which mediate transport across plasma membranes may be stored in secretory vesicles of absorptive or secretory epithelial cells. These vesicles are inserted into plasma membranes by exocytosis after appropriate physiological stimulation. We inquired if exocytosis of the abundant population of sub-apical vesicles in CAL+DCT cells is related to the PTH-stimulated increase of $[Ca^{2+}]_i$. Addition of bPTH(1–34), but not of bPTH(3–34) (a competitive inhibitor of PTH), caused a nearly four-fold increase in extracellular fluorescence of cells loaded with fluorescein-isothiocyanate-conjugated horseradish peroxidase (FITC-HRP) (Fig. 4a). This increase in extracellular fluorescence indicates that the contents of HRP-loaded vesicles were released by exocytosis after stimulation with PTH. As delivery of secretory vesicles to apical plasma membranes of epithelial cells requires functional microtubules^{25–27}, we tested the effect of the microtubule depolymerizing agent colchicine²⁸. Colchicine significantly blocked HRP release (Fig. 4a). It also abolished the stimulatory effects of PTH and of PTH+Bay k 8644 on $[Ca^{2+}]_i$ (Fig. 4b) but had no effect of its own on $[Ca^{2+}]_i$. This result is consistent with the view that PTH-stimulated exocytosis is necessary for initiation of Ca^{2+} entry. The incomplete inhibition of PTH-induced exocytosis of FITC-HRP may indicate the presence of a microtubule-independent component of exocytosis²⁶. These inhibitory actions of colchicine were not referable to a toxic effect nor to interference with PTH binding to its receptor, as colchicine did not interfere with the normal stimulation of cyclic AMP formation evoked by PTH (Fig. 4c).

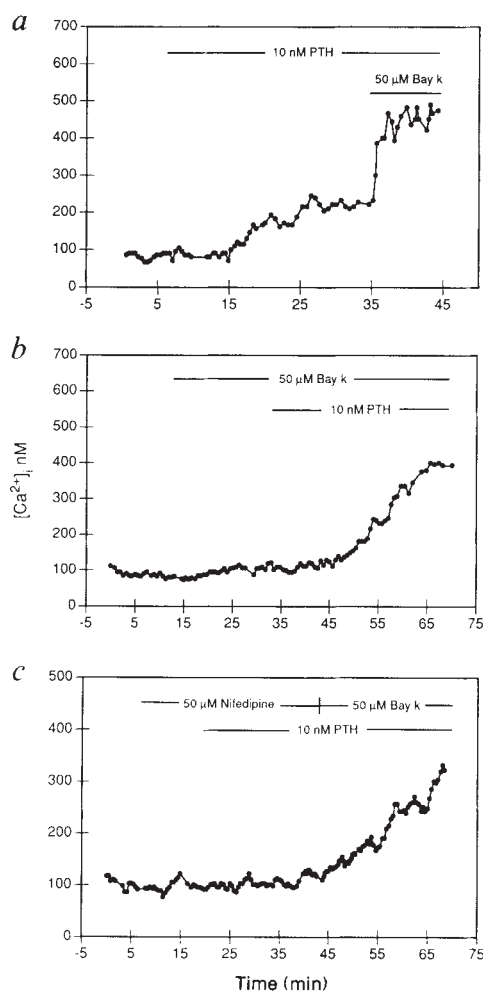


FIG. 3 Effects of dihydropyridine derivatives on $[Ca^{2+}]_i$. *a*, After PTH increased $[Ca^{2+}]_i$ to steady-state levels ($t=35$ min) addition of 50 μM racemic Bay k 8644 led to an immediate increase in $[Ca^{2+}]_i$. *b*, 50 μM Bay k 8644 had little or no effect of its own on $[Ca^{2+}]_i$. But, subsequent addition of 10 nM bPTH(1–34) increased $[Ca^{2+}]_i$ to about 450 nM. *c*, The dihydropyridine Ca^{2+} channel antagonist nifedipine (50 μM) had no effect on $[Ca^{2+}]_i$ but blocked the increase in $[Ca^{2+}]_i$ normally evoked by PTH. $[Ca^{2+}]_i$ increased promptly following equimolar replacement of nifedipine with Bay k 8644 at $t=45$ min, in the continued presence of PTH. Results of representative, single experiments are presented for clarity. Average values for $[Ca^{2+}]_i$ in each condition were as follows: control, 96 ± 5 ($n=17$); PTH, 260 ± 14 ($n=5$); PTH+Bay k, 478 ± 103 ($n=5$); Bay k, 103 ± 15 ($n=5$); nifedipine, 124 ± 9 ($n=8$). Separate experiments on Madin-Darby canine kidney cells (MDCK), a continuous line of epithelial cells expressing distal tubule phenotypic characteristics³⁵, indicate an apparent Michaelis constant K_m for activation of Ca^{2+} channels by Bay k 8644 after challenge with PTH of about 1.2 μM. (J. L. Flanagan and P.A.F., unpublished observations).

Our data provide evidence that PTH leads to the activation of dihydropyridine-sensitive Ca^{2+} channels in CAL+DCT kidney cells. The increase in $[\text{Ca}^{2+}]_i$ evoked by PTH represents an influx of extracellular Ca^{2+} . Once inserted or partially activated, the channels can be fully activated by Bay k 8644 or inhibited by nifedipine. These Ca^{2+} channels, or a factor leading

to their activation, seem to be stored in intracellular vesicles. On mobilization by PTH, these vesicles, which require microtubules for coherent delivery, fuse with plasma membranes. The channels, once inserted or activated, provide the uptake mechanism for Ca^{2+} into these Ca^{2+} -transporting epithelial cells. $\square$

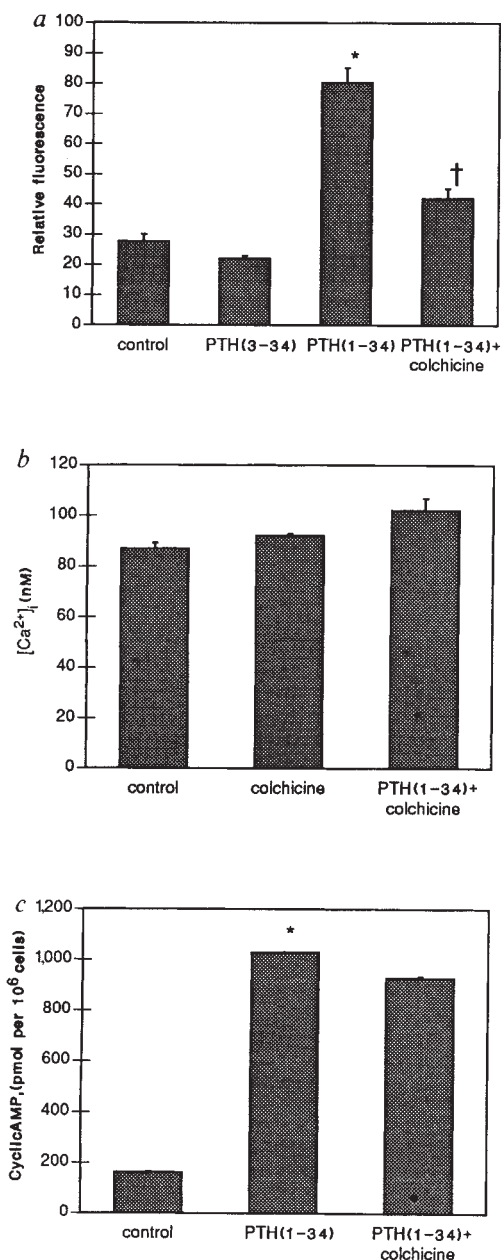


FIG. 4 Effects of colchicine on PTH-stimulated exocytosis (a), $[\text{Ca}^{2+}]_i$ (b), and formation of cAMP (c). a, Confluent monolayers of CAL + DCT cells were loaded with FITC-conjugated horseradish peroxidase (HRP; 1 mg ml^{-1}) by overnight incubation, a sufficient time for this fluid phase marker to equilibrate with intracellular vesicles³⁶. After washing, cells were treated with vehicle (control), 100 nM bPTH(3-34), 100 nM bPTH(1-34), or with bPTH(1-34) + $1 \mu\text{M}$ colchicine for 30 min. Extracellular medium was collected and fluorescence measured. Paired statistical comparisons were performed with Student's *t*-test. (* $P < 0.01$ versus control; † $P < 0.01$ versus PTH(1-34)). Absence of s.e.m. bars denotes inclusion within data point. b, $[\text{Ca}^{2+}]_i$ was measured as described in Fig. 1. Cells grown on coverslips were washed and pretreated for 5–10 min with $1 \mu\text{M}$ colchicine. Subsequent addition of 10 nM bPTH(1-34) had no effect on $[\text{Ca}^{2+}]_i$ at 25–30 min. c, Effect of colchicine on PTH-induced cAMP formation. Cyclic AMP was measured after a 15 min incubation with bPTH(1-34)³⁰. Addition of $1 \mu\text{M}$ bPTH(1-34) alone elicited a 560% increase in cAMP formation and was not inhibited by colchicine.

Received 18 May; accepted 6 July 1990.

1. Taylor, A. & Windhager, E. E. in *The Kidney: Physiology and Pathophysiology* (eds Seldin, D. W. & Giebisch, G.) 1297–1322 (Raven, New York, 1985).
2. Villereal, M. L. & Palfrey, H. C. A. *Rev. Nutr.* **9**, 347–376 (1989).
3. Pfaffner, W., Gstraunthaler, G., Kersting, U. & Oberleithner, H. *Renal Physiol. Biochem.* **12**, 328–337 (1989).
4. Wasserman, R. H. & Fullmer, C. S. *Adv. exp. Med. Biol.* **249**, 45–65 (1989).
5. Borke, J. L., Caride, A. J., Yaksh, T. L., Penniston, J. T. & Kumar, R. *Brain Res.* **489**, 355–360 (1989).
6. Borke, J. L. *et al.* *Am. J. Physiol.* **257**, C341–C346 (1989).
7. Trautwein, W. & Hescheler, J. A. *Rev. Physiol.* **52**, 257–274 (1990).
8. Hosey, M. M. & Lazunski, M. *J. Membrane Biol.* **104**, 81–105 (1988).
9. Friedman, P. A. *News Physiol. Sci.* **3**, 17–21 (1988).
10. Bronner, F. *Am. J. Physiol.* **257**, F707–F711 (1989).
11. van Os, C. H. *Prog. clin. biol. Res.* **252**, 223–229 (1988).
12. Borke, J. L., Penniston, J. T. & Kumar, R. *Semin. Nephrol.* **10**, 15–23 (1990).
13. Hruska, K. A. *et al.* *J. clin. Invest.* **79**, 230–239 (1987).
14. Goligorsky, M. S., Loftus, D. J. & Hruska, K. A. *Am. J. Physiol.* **251**, F938–F944 (1986).
15. Yamaguchi, D. T., Hahn, T. J., Iida-Klein, A., Kleeman, C. R. & Muallem, S. *J. biol. Chem.* **262**, 7711–7718 (1987).
16. Exton, J. H. in *Calcium-Binding Proteins in Health and Disease* (eds Norman, A. W., Vanaman, T. C. & Means, A. R.) 137–145 (Academic, New York, 1987).
17. Putney, J. W. Jr *Trends pharmac. Sci.* **8**, 481–486 (1987).
18. Hurwitz, L. A. *Rev. Pharmac. Tox.* **26**, 225–258 (1986).
19. Zeidel, M. L., Kikeri, D., Silva, P., Burrows, M. & Brenner, B. M. *J. clin. Invest.* **82**, 1067–1074 (1988).
20. McKenna, E., Koch, W. J., Sligh, D. F. & Schwartz, A. *Biochem. Pharmac.* **39**, 1145–1150 (1990).
21. Brown, D. *Am. J. Physiol.* **256**, F1–F12 (1989).
22. Wade, J. B. *Curr. Top. Membranes Transp.* **13**, 123–147 (1980).
23. Gluck, S., Cannon, C. & Al-Awqati, Q. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4327–4331 (1982).
24. Cannon, C., van Adelsberg, J., Kelly, S. & Al-Awqati, Q. *Nature* **314**, 443–446 (1985).
25. Achler, C., Filmer, D., Merte, C. & Drenckhahn, D. *J. Cell Biol.* **109**, 179–189 (1989).
26. Parczyk, K., Haase, W. & Kondor-Koch, C. *J. biol. Chem.* **264**, 16837–16846 (1989).
27. Molitoris, B. A. & Nelson, W. J. *J. clin. Invest.* **85**, 3–9 (1990).
28. Margolis, R. L. & Wilson, L. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3466–3470 (1977).
29. Allen, M. L. *et al.* *Am. J. Physiol.* **255**, F704–F710 (1988).
30. Friedman, P. A., Coutermarsh, B. A., Kennedy, S. M. & Pizzonia, J. H. *J. Bone Min. Res.* **4**, S346 (1989).
31. Grynkiewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440–3450 (1985).
32. Friedman, P. A. & Andreoli, T. E. *J. gen. Physiol.* **80**, 683–711 (1982).
33. Hebert, S. C., Friedman, P. A. & Andreoli, T. E. *J. Membrane Biol.* **80**, 201–219 (1984).
34. Friedman, P. A. *Am. J. Physiol.* **254**, F62–F70 (1988).
35. Handler, J. S. *J. exp. Biol.* **106**, 55–69 (1983).
36. Ajjoka, R. S. & Kaplan, J. *J. Cell Biol.* **104**, 77–85 (1987).

ACKNOWLEDGEMENTS. Bay k 8644 was supplied by A. Scriabine, Miles Laboratories, West Haven, Connecticut. We thank B. A. Stanton for thoughtful comments. These studies were performed in partial fulfillment of the requirements for a doctoral degree at Dartmouth College, Thayer School of Engineering (B.J.B.) and were supported by the NIH, American heart Association, and a Pharmaceutical Manufacturers Association Foundation Advanced Predoctoral Fellowship (B.J.B.).

Graded changes in dose of a *Xenopus* activin A homologue elicit stepwise transitions in embryonic cell fate

Jeremy B. A. Green & J. C. Smith

Laboratory of Embryogenesis, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

THE protein XTC-MIF, a *Xenopus* homologue of activin A and a potent mesoderm-inducing factor¹, can induce responding animal pole explants to form several different cell types in a dose-dependent manner, higher doses eliciting more dorso-anterior tissues². This graded response, characteristic of classically postulated morphogens³, may underlie pattern formation, but the response of intact animal caps to XTC-MIF provides only a crude indication of trends. Here we report the effects of XTC-MIF on dispersed blastomeres rather than intact animal caps. Under these conditions, responding cells distinguish sharply between doses of pure XTC-MIF differing by less than 1.5-fold. Two different response thresholds have been found, defining three cell states. This suggests that XTC-MIF has an instructive effect. Notochord and muscle are both induced in the same narrow dose-range. Mixing treated

with untreated cells does not seem to shift the dose thresholds, showing that at least some cells can stably record the received dose of inducing factor.

One source of heterogeneity in the previously reported response to XTC-MIF (ref. 2) probably derives from the structure of the responding animal caps. These are several cell layers thick and covered on one surface by a pigmented layer which is impervious to inducing factors⁴. It is possible to remove the pigmented layer of cells and to separate the remaining blastomeres into a single cell suspension by incubating caps in medium lacking calcium and magnesium. Such dispersed cells all differentiate as epidermis⁵. This fate can be suppressed by XTC-MIF, but to obtain the full range of differentiated mesodermal cell types (in the absence of additional exogenous proteins such as laminin and fibronectin⁶) it is necessary to reaggregate the cells after treatment⁵. We therefore used the strategy shown in Fig. 1; the results were analysed both by RNase protection and by immunofluorescence.

The results obtained with dispersed-reaggregated blastomeres were very different from those seen with intact animal caps². First, using RNase protection, we found that increasing the concentration of XTC-MIF by a factor of only 1.5 diverted cell aggregates from producing epidermal keratin messenger RNA to producing muscle-specific actin mRNA with no detectable areas of overlap (Fig. 2a). This threshold effect fits well with the idea that factor gradients can generate patterns with sharp boundaries and implies that such gradients do not have to be

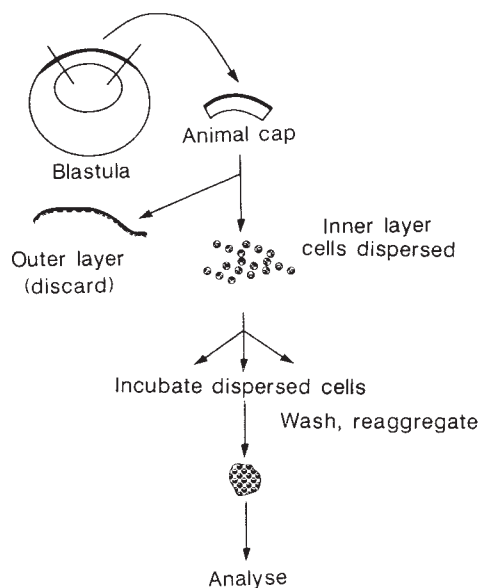


FIG. 1 Exposure of dispersed *Xenopus* animal pole blastomeres to XTC-MIF. Embryo preparation and media were as in ref. 5. Animal caps were excised from mid-blastula *Xenopus* embryos (stages 7.5–8.5) into calcium- and magnesium-free medium⁵ (CMFM). Inner layer cells could be disaggregated whereas the outer layers generally remained intact and were removed. After disaggregation, the inner layer cells were transferred to CMFM containing 0.1% BSA and appropriate concentrations of XTC-MIF in agarose-lined dishes. This operation was performed using glass Pasteur pipettes that had been treated briefly with 2% polyHEMA (Hydron Laboratories, New Brunswick, New Jersey) in ethanol/acetone (1:1) and allowed to dry. Cells were kept well dispersed by gentle pipetting. After one hour's incubation, cells were transferred to polyHEMA-treated microfuge tubes and centrifuged in a swing-out rotor at 165g for 5 min. Cells were washed once by resuspension in CMFM (resulting in an effective 100-fold dilution), recentrifugation and resuspension in 75% NAM (a further 100-fold dilution) before a final centrifugation. This procedure was always completed before the onset of gastrulation (stage 10). Pelleted cells were allowed to reaggregate overnight at 18 °C and incubated further in agarose-lined dishes until the equivalent of the tailbud stage (stage 18) for RNA analysis or tadpole stage (stage 30–35) for antibody staining.

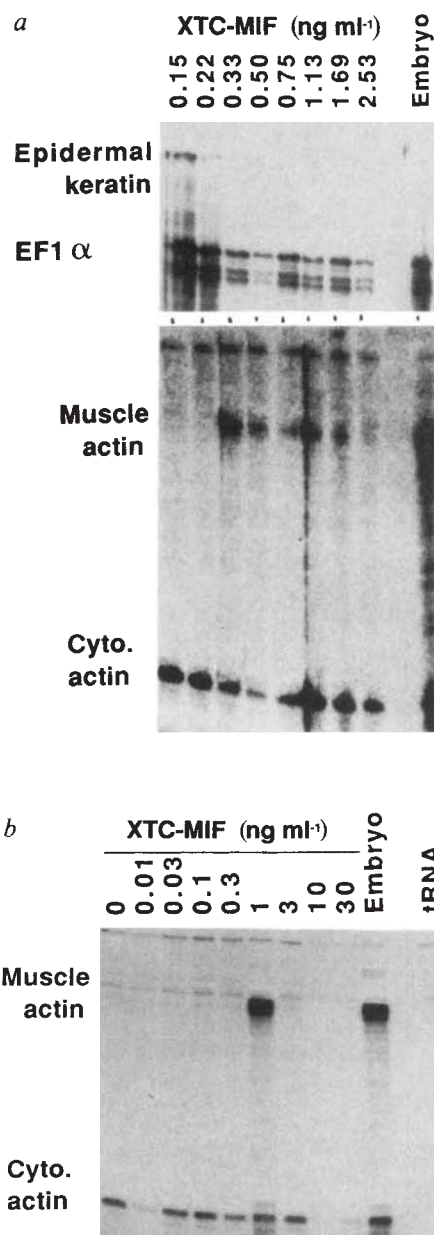


FIG. 2 Dose-response patterns for epidermal keratin and muscle actin in aggregates of dispersed blastomeres treated with XTC-MIF. Procedure essentially as Fig. 1. **a**, A 1.5-fold dilution series illustrates the sharp transition between epidermal (keratin-expressing) and muscle (cardiac actin-expressing) differentiation. **b**, A threefold dilution series demarcates the limits of the muscle-inducing XTC-MIF dose range. In this experiment, maximal muscle differentiation occurred at 1 ng ml⁻¹ XTC-MIF; this figure varied from 0.25 ng ml⁻¹ to 1.0 ng ml⁻¹ in different experiments, depending on egg batch and the exact stage at which blastomeres were treated with the factor (see also ref. 2). The level of cardiac actin gene expression observed in reaggregated blastomeres is similar to that seen in intact animal caps, consistent with the observation that not all cells in the aggregates differentiate as muscle (see Fig. 3).

METHODS. Purified XTC-MIF (ref. 1) was used at various dilutions to treat inner layer cells from about 200 excised blastula animal caps. RNA was extracted from aggregates at control stage 18 and RNase protection assays performed exactly as in ref. 2. Probes were: M2 for actins¹⁹; XK81 for epidermal keratin²⁰ (complementary DNA was a gift from I. B. Dawid, NIH, and an RNA probe was made by removing sequences downstream of the *Nco*I site, cloning the rest into a transcription vector, and linearizing at a vector site just upstream of the insert); EF1 α , gel-loading control expressed in all cells²¹, was made from a genomic clone (gift from P. Krieg, Austin) cloned into a transcription vector, nucleotides 194–100 (up to a *Hinf*I site) being transcribed. Each lane was loaded with RNA from inner cells of the equivalent of about four animal caps.

very steep. Further experiments showed that muscle is induced only within a narrow XTC-MIF concentration range: in Fig. 2b muscle-specific actin transcripts are observed at 1 ng ml^{-1} but not at 0.3 ng ml^{-1} or 3.0 ng ml^{-1} (see also Fig. 4, lanes 1–4). Using a muscle-specific antibody on sections of such aggregates, we confirmed this conclusion, but made two further observations. First, at concentrations near the edge of the muscle-inducing range we found aggregates with very small groups of muscle cells, containing perhaps as few as two nuclei (Fig. 3a, b). In terms of the 'community effect', where muscle differentiation requires the presence of several 'like-minded' cells⁷, this indicates that as few as two cells might constitute a community. Second, muscle-containing aggregates were conspicuously heterogeneous, with patches of notochord mixed with substantial patches of muscle, together with cells that bear neither marker (Fig. 3b). Previous results⁵ had indicated that notochord (the most 'dorsal' mesodermal tissue) is induced only at higher XTC-MIF doses. Co-induction of muscle and notochord demonstrates that there cannot be a simple correspondence between XTC-MIF concentration and dorsoventral position *in vivo* and that there must be other mechanisms generating such heterogeneity. One possibility is that the responding animal cap cells are heterogeneous, with only a subset of the cells able to make muscle (though at least 85% of cells can be diverted from the epidermal fate by XTC-MIF (ref. 5)), and indeed there is evidence for dorsoventral differences in the developmental potency of animal pole blastomeres^{8–10}. Alternatively, there may be interactions occurring after cells are reaggregated. These might include negative feedback ('regulation') on the proportions of cells that will differentiate as notochord or muscle¹¹, or

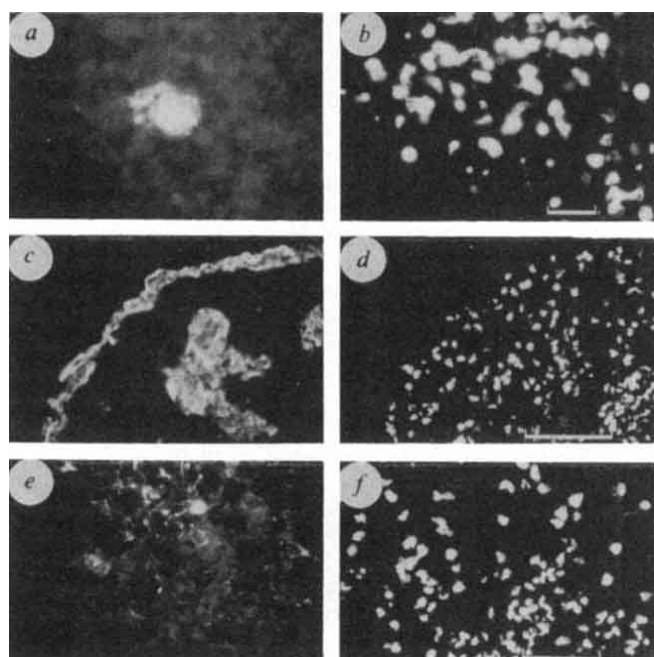


FIG. 3 Distribution of muscle and notochord in aggregates of dispersed blastomeres treated with XTC-MIF. a, Isolated cells labelled positively for muscle in aggregate of cells treated for 1 h with 0.125 ng ml^{-1} XTC-MIF. b, Nuclear staining of section shown in a. Scale bar (also applies to a), $30 \mu\text{m}$. c, Extensive but patchy muscle in aggregate of cells all treated for 1 h with 0.25 ng ml^{-1} XTC-MIF. d, Nuclear stain of c. Scale bar (also applies to c), $200 \mu\text{m}$. e, Notochord labelling in a nearby section of the aggregate shown in c. The labelling has a 'marbled' or 'veined' appearance because the antigen is specific to the extracellular notochord sheath²². f, Nuclear staining of e. Scale bar (also applies to e), $75 \mu\text{m}$.

METHODS. Aggregates and controls were fixed in 2% trichloroacetic acid at 4°C overnight. Other methods were exactly as in ref. 2. Antibodies were mouse monoclonal antibodies 12/101 for muscle²³ and MZ15 for notochord²².

even a positive 'autocatalytic' or 'community' effect. In this connection, it is significant that muscle and notochord tend to differentiate in patches and, furthermore, that notochord is often found in association with muscle. The role of cell interactions in groups of animal pole blastomeres after reaggregation is being investigated.

As we could find no direct marker for the tissue type induced by high concentrations of XTC-MIF, we mixed XTC-MIF-treated cells with untreated cells to detect indirect effects such as 'secondary inductions'. Treatment of dispersed cells with a range of concentrations of XTC-MIF (Fig. 4, lanes 1–4) results, after reaggregation, in little expression of the neural marker NCAM (neural cell adhesion molecule); the low level that is observed probably represents maternal transcripts¹². When identically treated cells were mixed with uninduced cells from the same egg batch (lanes 5–8), strong NCAM expression was observed, but only in the combinations containing blastomeres treated with concentrations of XTC-MIF too high to induce muscle-specific actin gene expression. We suggest that the NCAM is made by the originally uninduced component of the mixture, a conclusion based on embryological experiments showing that neural tissue arises from ectoderm through a secondary induction by mesoderm (reviewed in ref. 13), and one that we are attempting to test using cell lineage analysis and neural-

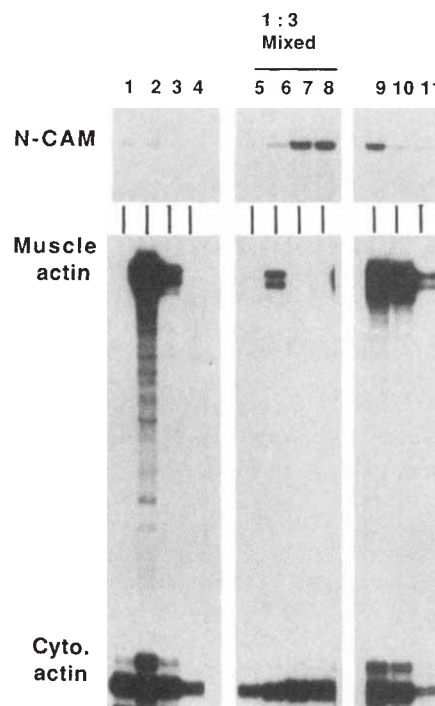


FIG. 4 Effect of mixing cells treated with XTC-MIF with untreated cells on muscle and neural (NCAM) expression in aggregates. Dispersed blastomeres were divided into two groups, one of which was treated for 1 h with various concentrations of XTC-MIF. Some of the treated cells were reaggregated without mixing (lanes 1–4), the remainder of the treated cells were mixed and reaggregated with at least a three-fold excess of untreated cells (lanes 5–8). XTC-MIF was from a batch assayed for biological activity rather than protein content. Concentrations were 1 unit (lanes 1, 5), 3 units (lanes 2, 6), 9 units (lanes 3, 7) and 27 units (lanes 4, 8), where 1 unit ml^{-1} is defined as the minimum concentration giving visible induction of an animal cap ($\sim 0.1 \text{ ng ml}^{-1}$ for XTC-MIF (ref. 2)). Lane 9, whole embryo; lanes 10 and 11, animal caps induced with 27 units and 3 units XTC-MIF, respectively. Aggregates were incubated until stage 18 and analysed by RNase protection. Muscle probe as Fig. 2; NCAM probe (gift from P. Krieg) has been shown to be neural-specific at the stage analysed¹². Sample loading was normalized for total animal cap-equivalents, so the reduced levels of muscle-specific actin in lanes 6 and 7 are due predominantly, if not entirely, to a fourfold reduction in the number of induced cells by virtue of mixing.

specific markers. From the results of these experiments, we conclude that animal pole blastomeres treated with high concentrations of XTC-MIF produce a strong neural inducing signal. This suggests that such concentrations have a physiological role in the embryo. Figure 4 also shows that the previously observed thresholds for muscle induction are not changed by cell mixing; at least some of the cells in the muscle-inducing range still give rise to muscle; that is, they stably recorded the concentration of XTC-MIF they experienced before mixing. This result argues against models in which modes of cell differentiation are determined by the proportions of cells that respond to an inductive signal¹⁴.

We have shown that XTC-MIF exhibits two of the properties required of a classical morphogen³: sharp threshold effects and multiple stable responses. Together, these imply an instructive mechanism. These properties are shared by the *bicoid* protein of *Drosophila*¹⁵, and it is possible that the mechanism by which thresholds are generated, through cooperative binding at many sites upstream of target genes¹⁶, is also shared. However, the pathways by which the two morphogens act must differ, because the *bicoid* protein is an intracellular DNA-binding transcription factor¹⁶ whereas XTC-MIF acts extracellularly⁴. That there is activin activity in the *Xenopus* blastula¹⁷ suggests that our results are physiologically and morphogenetically relevant. The co-induction of notochord and muscle means that a gradient of XTC-MIF cannot account for dorsoventral patterning. Rather, it is more as if it is involved in patterning an oblique axis running from forebrain to anus, with high concentrations inducing pre-

chordal mesoderm (for which no markers are currently available, but which is probably involved in substantial neuralization of overlying tissue) and lower concentrations inducing trunk mesoderm and proctodeum. The expression of the posterior mesoderm marker *Xhox3* in response to XTC-MIF (ref. 18) is consistent with this idea¹³. □

Received 17 May; accepted 31 July 1990.

1. Smith, J. C., Price, B. M. J., Van Nimmen, K. & Huylebroeck, D. *Nature* **345**, 729-731 (1990).
2. Green, J. B. A., Howes, G., Symes, K., Cooke, J. & Smith, J. C. *Development* **108**, 173-183 (1990).
3. Wolpert, L. *Development Suppl.* 3-12 (1989).
4. Cooke, J., Smith, J. C., Smith, E. J. & Yaqoob, M. *Development* **105**, 549-558 (1987).
5. Symes, K., Yaqoob, M. & Smith, J. C. *Development* **104**, 609-618 (1988).
6. Godsave, S. F. & Slack, J. M. W. *Dev Biol.* **134**, 486-490 (1989).
7. Gurdon, J. B. *Nature* **336**, 772-774 (1988).
8. Savage, R. & Phillips, C. R. *Dev Biol.* **133**, 157-168 (1989).
9. Sharpe, C. R., Fritz, A., De Robertis, E. M. & Gurdon, J. B. *Cell* **50**, 749-758 (1987).
10. Dixon, J. E. & Kintner, C. R. *Development* **106**, 749-757 (1989).
11. Cooke, J. *J. Embryol. exp. Morph.* **76**, 95-114 (1983).
12. Kintner, C. R. & Melton, D. A. *Development* **99**, 311-325 (1987).
13. Smith, J. C. *Curr. Op. Cell Biol.* **1**, 1061-1070 (1989).
14. Smith, J. C., Cooke, J., Green, J. B. A., Howes, G. & Symes, K. *Development Suppl.* 149-159 (1989).
15. Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 95-104 (1988).
16. Driever, W. & Nüsslein-Volhard, C. *Nature* **337**, 138-143 (1989).
17. van den Eijnden-Van Raaij, et al. *Nature* **345**, 732-734 (1990).
18. Ruiz-i-Altaba, A. & Melton, D. A. *Nature* **341**, 33-38 (1989).
19. Mohun, T. J., Brennan, S., Dathan, N., Fairman, S. & Gurdon, J. B. *Nature* **311**, 716-721 (1984).
20. Jonas, E., Sargent, T. D. & Dawid, I. B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5413-5417 (1985).
21. Krieg, P. A., Varum, S. M., Wormington, W. M. & Melton, D. A. *Dev Biol.* **133**, 93-100 (1989).
22. Smith, J. C. & Watt, F. M. *Differentiation* **29**, 109-115 (1985).
23. Kintner, C. R. & Brockes, J. P. *Nature* **308**, 67-69 (1984).

ACKNOWLEDGEMENTS. We thank our colleagues M. Sargent and J. Cooke for helpful discussion and critical reading of the manuscript. This work was supported by the MRC. J.B.A.G. is an MRC Training Fellow.

Downregulation of enkephalin-mediated inflammatory responses by CD10/neutral endopeptidase 24.11

Margaret A. Shipp^{*†}, George B. Stefano[‡], Luciano D'Adamo^{*§}, Stephanie N. Switzer^{*}, Frank D. Howard^{*†}, Juan Sinisterra[‡], Berta Scharrer[¶] & Ellis L. Reinherz^{*†}

^{*}Laboratory of Immunobiology, Dana-Farber Cancer Institute and Departments of [†]Medicine and [§]Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

[‡]Multidisciplinary Center for the Study of Aging, State University of New York, College at Old Westbury, Old Westbury, New York 11568, USA

[¶]Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA

THE antigen CD10 (common acute lymphoblastic leukaemia antigen), which is the zinc metalloprotease, neutral endopeptidase 24.11 (also known as NEP or 'enkephalinase')¹⁻⁷, is expressed by acute lymphoblastic leukaemias^{8,9}, normal lymphoid progenitors¹⁰, mature polymorphonuclear leukocytes¹¹ and certain non-haematopoietic cells¹². CD10/NEP hydrolyses several naturally occurring peptides, including the endogenous opioid pentapeptides Met- and Leu-enkephalin⁶. In invertebrate organisms such as the mollusc *Mytilus edulis*, Met-enkephalin triggers inflammatory responses by inducing morphological changes, directed migration and aggregation of haemocytes^{13,14}. We report here that a structure related to CD10/NEP is expressed by *M. edulis* haemocytes and that abrogation of CD10/NEP enzymatic activity reduces the amount of Met-enkephalin required for haemocyte activation by five orders of magnitude. Similar results are obtained with CD10⁺ human polymorphonuclear leukocytes, indicating that CD10/NEP related structures regulate enkephalin-mediated inflammatory

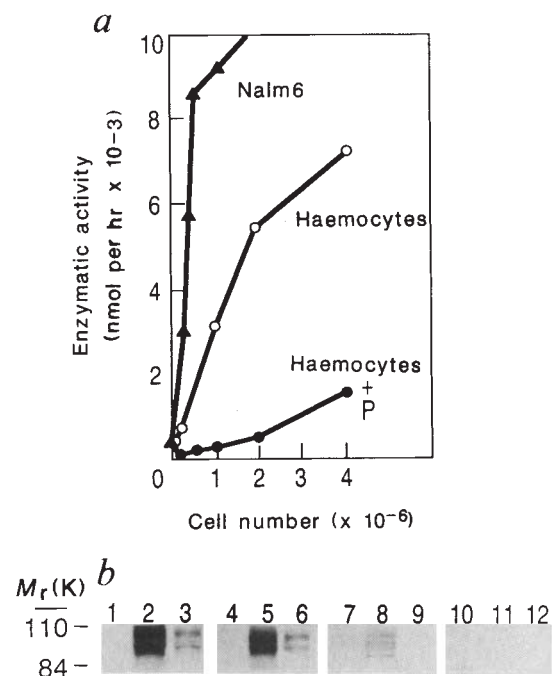
responses in organisms whose ancestors diverged ~500 million years ago¹⁵.

Whole cell suspensions of *M. edulis* haemocytes were analysed for CD10/NEP enzymatic activity² (Fig. 1a). Like human CD10⁺ Nalm-6 cells, *M. edulis* haemocytes have readily detectable CD10/NEP activity (~3,000 nmol per h per 10⁶ cells), which is markedly reduced by the addition of the specific NEP inhibitor phosphoramidon¹⁶. To characterize the invertebrate enzyme further, *M. edulis* haemocyte lysates were size-fractionated, blotted and probed with a heteroantiserum (anti-human (hu) CD10) raised against denatured, gel-purified human CD10 protein (Fig. 1b). Anti-hu-CD10 identifies a series of proteins of relative molecular mass 95,000-100,000 (*M_r* 95-100 K) in CD10⁺ human cell lines such as Nalm-6 and Laz 221 that are not found in the CD10⁻ cell line Laz 388 (Fig. 1b, lanes 2, 5, 11) and presumably represent differentially processed forms of the metalloprotease. A similarly sized set of proteins (Fig. 1b, lane 8) is specifically detected in *M. edulis* haemocytes; the addition of denatured purified CD10/NEP eliminates anti-hu-CD10 reactivity (Fig. 1c, lane 9). Although absolute proof of structural homology between invertebrate and vertebrate CD10/NEP molecules requires sequence information, both the immunoreactivity and functional data are consistent with structural relatedness.

To determine whether the putative invertebrate CD10/NEP regulates local enkephalin levels and thereby modulates enkephalin-mediated changes in cell shape, migration and aggregation, *M. edulis* haemocytes were stimulated with Met-enkephalin at 10⁻⁶-10⁻¹² M in the presence or absence of three chemically unrelated NEP inhibitors, phosphoramidon, SCH32615¹⁷ or thiorphan¹⁶. Phosphoramidon reduces the minimum concentration of Met-enkephalin required to trigger haemocyte morphological changes and increases in cell area by five orders of magnitude (Fig. 2, upper left panel, upper right panel (lanes 1 and 2)). The minimum concentration of Met-enkephalin required to activate haemocytes is similarly reduced by the unrelated specific NEP inhibitor, SCH32615, and by thiorphan, which inhibits NEP and peptidyl dipeptidase with

FIG. 1 Analysis of CD10/NEP enzymatic activity and protein in *M. edulis* haemocytes. **a**, CD10/NEP enzymatic activity in whole cell suspensions. The levels of CD10/NEP enzymatic activity associated with increasing numbers of *M. edulis* haemocytes (○-○) or Nalm-6 cells (▲-▲) are shown and compared with CD10/NEP enzymatic activity of *M. edulis* haemocytes treated with 20 μ M phosphoramidon (*N*-[α -L-rhamnopyranosyloxyhydroxyphosphinyl]-L-leucyl-L-tryptophan) (●-●). **b**, Western blot analysis of *M. edulis* haemocyte CD10/NEP. Equivalent concentrations of cell lysate from the CD10⁺ human lymphoblastic leukaemia cell lines Nalm-6 (lanes 1-3) and Laz 221 (lanes 4-6), *M. edulis* haemocytes (lanes 7-9) and the CD10⁻ human EBV-transformed lymphoblastoid line Laz 388 (lanes 10-12) were size-fractionated, blotted and probed with either preimmune sera (lanes 1, 4, 7, 10), the heterosera directed against denatured human CD10, anti-hu-CD10 (lanes 2, 5, 8, 11), or anti-hu-CD10 preincubated with 50 μ g of denatured gel-purified CD10/NEP (lanes 3, 6, 9, 12).

METHODS. CD10/NEP activity was measured fluorometrically in a coupled assay using glutaryl-Ala-Ala-Phe-4-methoxy-2-naphthylamide as substrate². The assay was performed as previously described² except that Nalm-6 cells and haemocytes were resuspended in hypertonic media (Instant Ocean; NaCl 480 mM, KCl 10 mM, CaCl₂ 10 mM, MgCl₂ 20 mM, MgSO₄ 30 mM, HEPES buffer 5 mM) for analysis. In certain experiments, 20 μ M phosphoramidon was added to the cell suspension immediately before adding the substrate. Nalm-6, Laz 221 and Laz 388 cells and *M. edulis* haemocyte cell lysates were size-fractionated in 7.5% SDS-PAGE reducing gels and transferred to nitrocellulose. Blots were incubated sequentially with antisera and goat anti-rabbit immunoglobulin alkaline phosphatase (Biorad), and developed with 5-bromo-4-chloro-3-indoyl phosphate and nitroblue tetrazolium (Biorad). The anti-hu-CD10 heterosera was generated by immunizing rabbits with human denatured CD10 which was purified from Nalm-6 cells as before¹. Subtidal *M. edulis* were collected and haemocytes isolated as previously described¹³. Haemocytes were washed extensively in 4 °C hypertonic media and used immediately thereafter. The EBV-transformed B lymphoblastoid



line, Laz 388, was derived from normal lymphocytes from the donor of the acute lymphoblastoid leukaemia cell line, Laz 221.

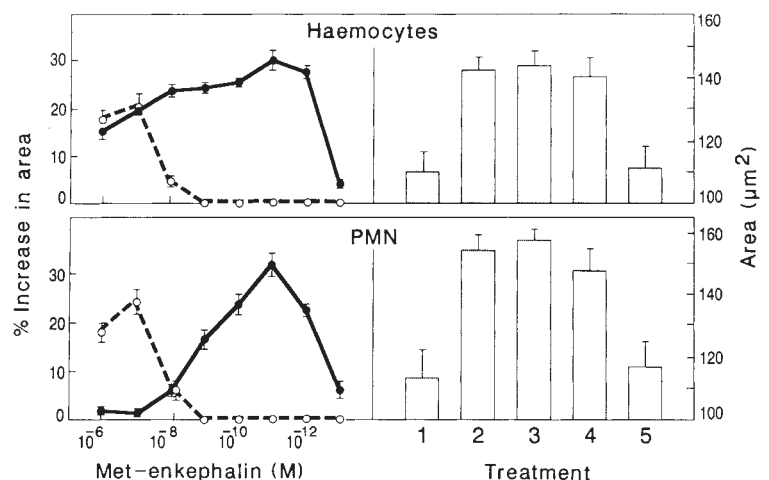
similar efficiency (Fig. 2, upper right panel (lanes 3 and 4))¹⁶. In contrast, captopril, which inhibits peptidyl dipeptidase but is inactive against NEP¹⁶, has no effect on haemocyte activation triggered by Met-enkephalin (Fig. 2, upper right panel (lane 5)). Therefore, the observed reduction in Met-enkephalin concentration required for haemocyte activation results from specific inhibition of a CD10/NEP related structure.

An unidentified subpopulation of human peripheral blood cells responds to Met-enkephalin stimulation in a manner similar to that of *M. edulis* haemocytes¹⁴. Expression of CD10/NEP

on human polymorphonuclear leukocytes¹¹ prompted us to determine whether these cells respond to Met-enkephalin and whether CD10/NEP affects Met-enkephalin stimulation. Purified subpopulations of human polymorphonuclear leukocytes, lymphocytes and monocytes were stimulated with Met-enkephalin at 10^{-6} – 10^{-12} in the presence or absence of NEP inhibitors. Although Met-enkephalin had no detectable effect on peripheral blood lymphocytes or monocytes during the observation period (data not shown), the neuropeptide induced changes in the cell area of polymorphonuclear

FIG. 2 Met-enkephalin stimulation of *M. edulis* haemocytes and human polymorphonuclear leukocytes is enhanced by CD10/NEP inhibition. Left upper and lower panels, increases in cell area of *M. edulis* haemocytes (upper panel) or human polymorphonuclear leukocytes (lower panel) stimulated with 10^{-6} – 10^{-13} M Met-enkephalin in the presence (●-●) or absence (○-○) of phosphoramidon. Results are displayed as mean $\pm$ s.d. Right upper and lower panels, increases in cell area of *M. edulis* haemocytes (upper panel) or human polymorphonuclear leukocytes (lower panel) stimulated with Met-enkephalin at 10^{-11} M either alone (lane 1), or Met-enkephalin at 10^{-11} M plus either phosphoramidon (lane 2), SCH32615 (lane 3), thiorphan (lane 4) or captopril (lane 5). Results are displayed as mean $\pm$ s.d.

METHODS. Haemocytes in 0.1 ml haemolymph were mixed with 0.1 ml of hypertonic media (Instant Ocean) and Met-enkephalin (Penninsula Laboratories) at the specified concentration at room temperature. In certain experiments, one of three CD10/NEP inhibitors, phosphoramidon (Sigma), thiorphan (*N*-[DL-2-benzyl-3-mercaptopropionyl] glycine; Sigma) or SCH32615 (*N*-[L-(-1-carboxy-2-phenyl)ethyl]-L-phenylalanyl- β -alanine; gift from Schering-Plough, Bloomfield, NJ), or the peptidyl dipeptidase inhibitor captopril (D-3-mercapto-2-methyl propanoyl-L-proline; gift from L. Hersh, University of Texas Southwestern Medical Center, Dallas) was also added at 100 μ M to the cells. Cell suspensions were added to a coverslip precoated with 2% bovine serum albumin and analysed by time-lapse video recording using phase contrast and Nomarski optics on a Zeiss Axiophot microscope. Human polymorphonuclear leukocytes, monocytes and lym-



phocytes were obtained from the peripheral blood of volunteer donors, as previously described²⁵. Human peripheral blood subpopulations were resuspended in RPMI medium rather than hypertonic medium and stimulated at 37 °C rather than room temperature. The changes in shape and area of individual cells were evaluated at 5-min intervals for 15 min.

leukocytes that were similar to those seen in *M. edulis* haemocytes (Fig. 2, lower panel). In the presence of phosphoramidon, increases in the cell area of polymorphonuclear leukocytes occurred at reduced Met-enkephalin concentrations (10^{-9} – 10^{-12} M) (Fig. 2, lower left panel and lower right panel (lanes 1, 2)). The minimum concentration of Met-enkephalin required to activate polymorphonuclear leukocytes was similarly reduced by SCH32615 and thiorphan but was unaffected by captopril (Fig. 2, lower right panel (lanes 3–5)).

In Fig. 3a–d, the changes in polymorphonuclear leukocyte morphology associated with stimulation by Met-enkephalin in the presence or absence of phosphoramidon are shown. Polymorphonuclear leukocytes treated with 10^{-6} M Met-enkephalin assume a flattened amoeboid shape and develop pseudopodia, whereas polymorphonuclear leukocytes treated with Met-enkephalin (10^{-11} M) or phosphoramidon alone retain their rounded inactive shape (Fig. 3a–c). Polymorphonuclear leukocytes treated with 10^{-11} M Met-enkephalin in combination with phosphoramidon (Fig. 3d), however, undergo morphological changes similar to those seen in cells treated with 100,000-fold higher Met-enkephalin concentrations (Fig. 3a). Morphological changes such as those shown in Fig. 3a and d are associated with an increase in cell area of ~25% and precede migration and the development of cell clusters^{13,14}.

The regulation of enkephalin signals in haematopoietic cells by coexpression of cell surface opioid receptors and the CD10/NEP hydrolytic enzyme resembles the regulation of cholinergic signals at the neuromuscular junction where both acetylcholine receptors and acetylcholinesterase are expressed^{18,19}. Inhibition of acetylcholinesterase potentiates membrane depolarization following acetylcholine stimulation¹⁹ in the same way that inhibition of CD10/NEP augments the activation of haemocytes and polymorphonuclear leukocytes following enkephalin stimulation.

The important role of opioid neuropeptides in modulating inflammatory responses underscores the importance of activated T cells as a source of enkephalin production distinct from the nervous system. As up to 0.5% of the messenger RNA made by activated helper T cells encodes preproenkephalin and activated helper T cells secrete immunoreactive Met-enkephalin²⁰, enkephalins may function as primitive cytokines in both vertebrates and invertebrates.

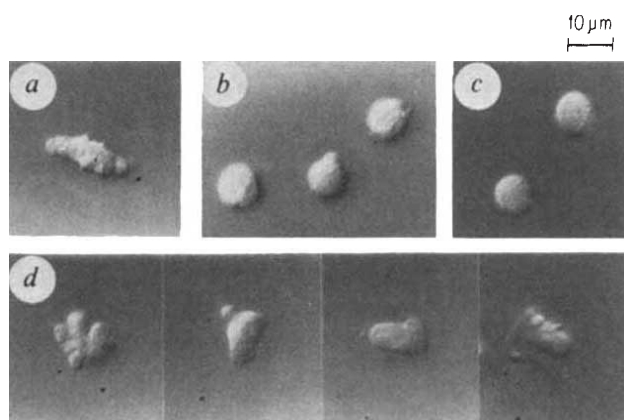


FIG. 3 Met-enkephalin stimulation of human polymorphonuclear leukocytes is enhanced by CD10/NEP inhibition. Human polymorphonuclear leukocytes were stimulated with Met-enkephalin (10^{-6} M) (a), Met-enkephalin (10^{-11} M) (b), phosphoramidon alone (c), or Met-enkephalin (10^{-11} M) in combination with phosphoramidon (100 μ M) (d). Changes in cell shape are quantified using the form-factor-PE equation, $ff = 4 \times \pi \times \text{area/perimeter}^2 \times 100$ (ref. 13). Inactive rounded cells with smaller perimeters have higher ff values, whereas activated cells with more amoeboid shapes and larger perimeters have lower ff values. The ff values for the indicated cells are: a, ff 45 ± 11 ; b, ff 86 ± 4.3 ; c, ff 85 ± 1 ; d, ff 45 ± 5.2 . Photomicrographs were obtained on a Zeiss Axiophot using Nomarski optics, a $\times 100$ objective, and a video printer.

Two additional CD10/NEP substrates, fMet-Leu-Phe (ref. 21) and substance P (ref. 22) stimulate inflammatory responses. fMet-Leu-Phe is the principal chemotactic peptide produced by *Escherichia coli* and the tachykinin substance P is a primary mediator of neurogenic inflammation of the respiratory tract²³. Therefore, it is likely that CD10/NEP also modulates fMet-Leu-Phe- and substance P-mediated inflammatory responses by controlling local peptide concentrations in haematopoietic cells and parenchymal tissues^{21,24}. Taken together, these results suggest that CD10/NEP functions in multiple organ systems, perhaps including lymphoid progenitors, to downregulate induced responses to peptide hormones and that the enzyme's role in regulating local peptide concentrations is of fundamental importance to both invertebrate and mammalian biology. □

Received 7 June; accepted 23 July 1990.

- Shipp, M. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4819–4823 (1988).
- Shipp, M. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 297–301 (1989).
- Letarte, M. *et al. J. exp. Med.* **168**, 1247–1253 (1988).
- Kerr, M. A. & Kenny, A. J. *Biochem. J.* **137**, 477–488 (1974).
- Kerr, M. A. & Kenny, A. J. *Biochem. J.* **137**, 489–495 (1974).
- Malfroy, B., Swertz, J. B., Guyon, A., Roques, B. P. & Schwartz, J. C. *Nature* **276**, 523–526 (1978).
- Malfroy, B. *et al. FEBS Lett.* **229**, 206–210 (1988).
- Brown, G., Hogg, N. & Greaves, M. *Nature* **258**, 454–456 (1975).
- Ritz, J., Pesando, J. M., Notis-McConarty, J., Lazarus, H. & Schlossman, S. F. *Nature* **283**, 583–585 (1980).
- Greaves, M. F. *et al. Blood* **61**, 628–639 (1983).
- Cossmann, J., Neckers, L. M., Leonard, W. J. & Greene, W. C. *J. exp. Med.* **157**, 1064–1069 (1983).
- Metzgar, R. S., Borowitz, M. J., Jones, N. H. & Dowell, B. L. *J. exp. Med.* **154**, 1249–1254 (1981).
- Stefano, G. B., Leung, M. K., Zhao, X. & Scharrer, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 626–630 (1989).
- Stefano, G. B., Cadet, P. & Scharrer, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6307–6311 (1989).
- Meglitsch, P. A. in *Invertebrate Zoology* (Oxford University Press, New York, 1967).
- Matsas, R., Kenny, J. & Turner, A. J. *Biochem. J.* **223**, 433–440 (1984).
- Martins, M. A., Shore, S. A., Gerard, N. P., Gerard, C. & Drazen, J. M. *J. clin. Invest.* **85**, 170–176 (1990).
- Raftery, M. A., Hunkapiller, M. W., Strader, C. D. & Hood, L. E. *Science* **208**, 1454–1456 (1980).
- Massoulié, J. & Bon, S. A. *Rev. Neurosci.* **5**, 57–106 (1982).
- Zurawski, G. *et al. Science* **232**, 772–775 (1986).
- Painter, R. G. *et al. J. biol. Chem.* **263**, 9456–9461 (1988).
- Kenny, A. J. *Biomed. biochim. Acta* **45**, 1503–1513 (1986).
- Lundberg, J. M., Saria, A., Brodin, S., Rosell, S. & Folkers, K. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1120–1124 (1983).
- Umeno, E., Nadel, J. A., Huang, H.-T. & MacDonald, D. M. *J. appl. Physiol.* **66**, 2647–2652 (1989).
- Boyum, A. *Scand. J. clin. Lab. Invest.* **21**, 77–89 (1968).

ACKNOWLEDGEMENTS. We thank J. Griffin and L. B. Chen for their review of the manuscript. This work was supported in part by the NIH.

Co-capping of *ras* proteins with surface immunoglobulins in B lymphocytes

Linda Graziadei, Karl Riabowol & Dafna Bar-Sagi*

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

CELLULAR *ras* genes encode a family of membrane-associated proteins ($p21^{ras}$) that bind guanine nucleotide and possess a low intrinsic GTPase activity^{1–3}. The $p21^{ras}$ proteins are ubiquitously expressed in mammalian cells and are thought to be involved in a growth-promoting signal transduction pathway⁴; their mode of action, however, remains unknown. The ligand-induced movement of cell-surface receptors seems to be a primary event in the transduction of several extracellular signals that control cell growth and differentiation. In B lymphocytes, surface immunoglobulin receptors crosslinked by antibody or other multivalent ligands form aggregates called patches, which then collect into a single assembly, a cap, at one pole of the cell^{5,6}. This process constitutes the initial signal for the activation of a B cell. Here we show by immunofluorescence microscopy that $p21^{ras}$ co-caps with surface immunoglobulin molecules in mouse splenic B lymphocytes. In contrast, no apparent change in the distribution of

* To whom correspondence should be addressed.

p21^{ras} occurs during the capping of concanavalin A receptors. The redistribution of p21^{ras} is apparent at the early stages (patching) of immunoglobulin capping and is inhibited by metabolic inhibitors and the cytoskeleton-disrupting agents colchicine and cytochalasin D. The distribution of another membrane-associated guanine nucleotide-binding regulatory protein, the G_i α subunit, is not affected by surface immunoglobulin capping. These findings demonstrate that p21^{ras} can migrate in a directed manner along the plasma membrane and suggest that p21^{ras} may be a component of the signalling pathway initiated by the capping of surface immunoglobulin in B lymphocytes.

To examine the distribution of p21^{ras} in B lymphocytes, we used affinity-purified antibodies raised against a bacterially expressed human Ha-ras protein⁷. The specificity of the antibodies was determined by immunoprecipitation and immunofluorescence assays. The antibody specifically recognized p21^{ras} polypeptides from spleen lymphocytes (Fig. 1A, lane 2), and from normal rat kidney cells transformed with v-Ki-ras (KNRK cells) (Fig. 1A, lane 4). Preincubation of the anti-p21^{ras} antibodies with an excess of purified antigen (Ha-ras) prevented the precipitation of the p21^{ras} polypeptide in both spleen lymphocytes (Fig. 1A, lane 1) and KNRK cells (Fig. 1A, lane 3). Indirect immunofluorescence experiments (Fig. 1B) showed that the staining was confined to the cell periphery both in KNRK cells (Fig. 1B, c) and in spleen lymphocytes (Fig. 1B, a). The anti-p21^{ras} antibody did not label living intact lymphocytes; that is, permeabilization of cells was required. Therefore, the immunofluorescent staining of p21^{ras} was intracellular, consistent with the known localization of p21^{ras} to the inner face of the plasma membrane⁸. This staining was abolished by pre-incubation of the antibody solution with an excess of purified Ha-ras protein (Fig. 1B, b and d).

Using double immunofluorescent labelling, the distributions of p21^{ras} and surface immunoglobulin were determined during

TABLE 1 Effects of metabolic inhibition and cytoskeletal disruption on the anti-IgG-induced redistribution of IgG and p21^{ras}

	IgG	% Capped p21 ^{ras}
Control	75 $\pm$ 5	76 $\pm$ 7
Sodium azide	6 $\pm$ 4	5 $\pm$ 3
Cytochalasin D* and colchicine†	0	1 $\pm$ 2

Spleen cells were pre-incubated with sodium azide (15 mM) or with cytochalasin D (40 μ M) plus colchicine (100 μ M) for 2 h at 37 °C. Cells were then treated with fluorescein-conjugated goat anti-mouse IgG (100 μ g ml⁻¹), in the presence of drugs, for 30 min at 20 °C. Fixation and indirect immunofluorescence staining were carried out as described in the legend to Fig. 1. Cells were examined for rhodamine and fluorescein staining, and the percentage of cells showing capped staining was determined. At least 200 cells were counted from each of two independent experiments. Values shown are mean $\pm$ s.d.

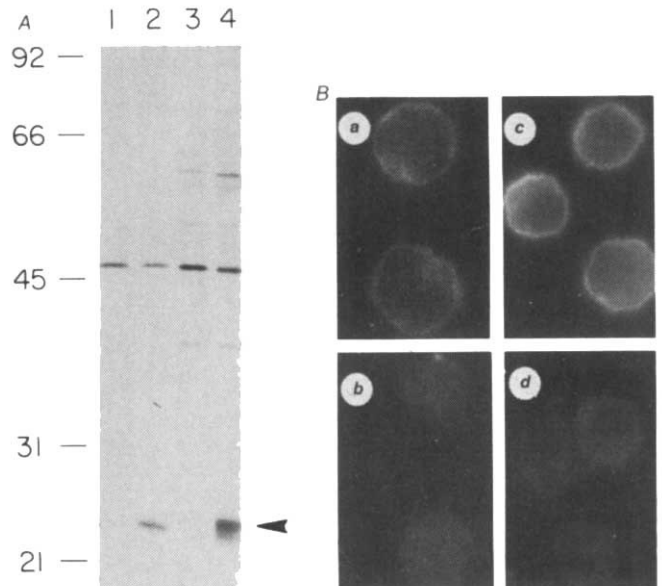
* Cytochalasin D was dissolved in dimethyl sulphoxide (DMSO). The final DMSO concentration was 2%. In a control experiment, DMSO alone had no visible effect on the cells.

† Colchicine alone (100 μ M) had no effect on capping.

the process of capping. Mouse spleen lymphocytes were incubated with fluorescein-conjugated goat anti-mouse immunoglobulin antibodies (GAMIg) at 20 °C for various intervals, fixed and then stained with rabbit anti-p21^{ras} antibodies followed by rhodamine-conjugated goat anti-rabbit IgG antibodies. In resting splenic lymphocytes the staining for both surface immunoglobulin and p21^{ras} was found to be uniformly distributed along the cell surface (Fig. 2A, a and b). When lymphocytes were stimulated to cap surface immunoglobulin by exposure to GAMIg antibodies for varying times, the p21^{ras} staining redistributed concomitantly with the capping of surface

FIG. 1 A, Immunoprecipitation of p21^{ras} protein by anti-p21^{ras} antibodies. Arrowhead marks the positions of p21^{ras} polypeptides. Lanes 1 and 2, immunoprecipitations from mouse splenic lymphocytes with anti-p21^{ras} antibodies that had been preincubated with (lane 1) or without (lane 2) purified antigen. Lanes 3 and 4, immunoprecipitations from KNRK cells with anti-p21^{ras} antibodies that had been preincubated with (lane 3) or without (lane 4) purified antigen. These cells typically exhibit a high level of p21^{ras} expression. The numbers in the left margin represent the migration position of relative molecular mass (M_r) standards ($M_r \times 10^{-3}$). B, Localization of p21^{ras} by immunofluorescence. Freshly isolated mouse splenic lymphocytes (a) and ras-transformed (KNRK) cells (c) were fixed in 3.7% formaldehyde and permeabilized by brief extraction with PBS containing 0.2% Triton X-100. The cells were stained with affinity-purified anti-p21^{ras} antibodies (100 μ g ml⁻¹, a and c) or with anti-p21^{ras} antibodies that had been blocked with an excess of purified c-Ha-ras (b and d, respectively), followed by rhodamine-conjugated affinity-purified goat anti-rabbit IgG antibodies (Cappel Research Reagents). (a and b, $\times 4,600$) (c and d, $\times 780$).

METHODS. Polyclonal rabbit antibodies were raised against purified amino-truncated human Ha-ras protein (lacking codons 1–23) produced in an *Escherichia coli* expression system⁷. Specific anti-p21^{ras} antibodies were purified by affinity chromatography using immobilized antigen. Lymphocytes were obtained from spleens of BALB/c mice by teasing the spleens in RPMI 1640 medium containing 2% FCS. The suspension was depleted of red cells by a 5-min incubation in 0.75% NH₄Cl at 4 °C. The cells were then rinsed three times with RPMI 1640, and the final pellet suspended in the labelling medium. Rat kidney fibroblasts transformed with Kirsten murine sarcoma virus (KNRK, ATCC CRL 1569) were grown in DMEM medium with 10% FCS. The cells were seeded in 6-cm dishes 2 days before addition of radioactive label, at which point they were actively growing (~60% confluent). Labelling was carried out for 3 h at 37 °C in methionine-free DMEM containing 5% methionine-free FCS and 500 μ Ci ml⁻¹ [³⁵S]methionine. The cells were collected, washed three times in PBS and resuspended and boiled for 2 min in lysis buffer containing 150 mM NaCl, 10 mM Tris (pH 7.4), 1% Triton X-100, 1% sodium deoxycholate, 1% sodium dodecyl sulphate (SDS), and 1 mM EDTA. Before immunoprecipitation, the concentration of SDS in lysates was adjusted to 0.1%, and lysates were clarified by centrifugation at 12,000g



at 4 °C for 10 min. The supernatants were reacted with the affinity-purified anti-p21^{ras} antibodies (2 μ g). For control precipitations, the anti-p21^{ras} antibodies were blocked by pre-incubation with 5 μ g antigen for 1 h at 4 °C. Each immunoprecipitation reaction contained 5×10^7 c.p.m. and was carried out for 3 h at 4 °C. The immune complexes were collected using a 50% solution of protein A-Sepharose, washed five times in lysis buffer containing 0.1% SDS, then solubilized in SDS-PAGE sample buffer. The immunoprecipitated material was analysed on a 12.5% SDS-polyacrylamide gel. The gel was impregnated with ENHANCE (NEN Research Products), dried and exposed for 1.5 days to XAR film (Kodak) at -70 °C.

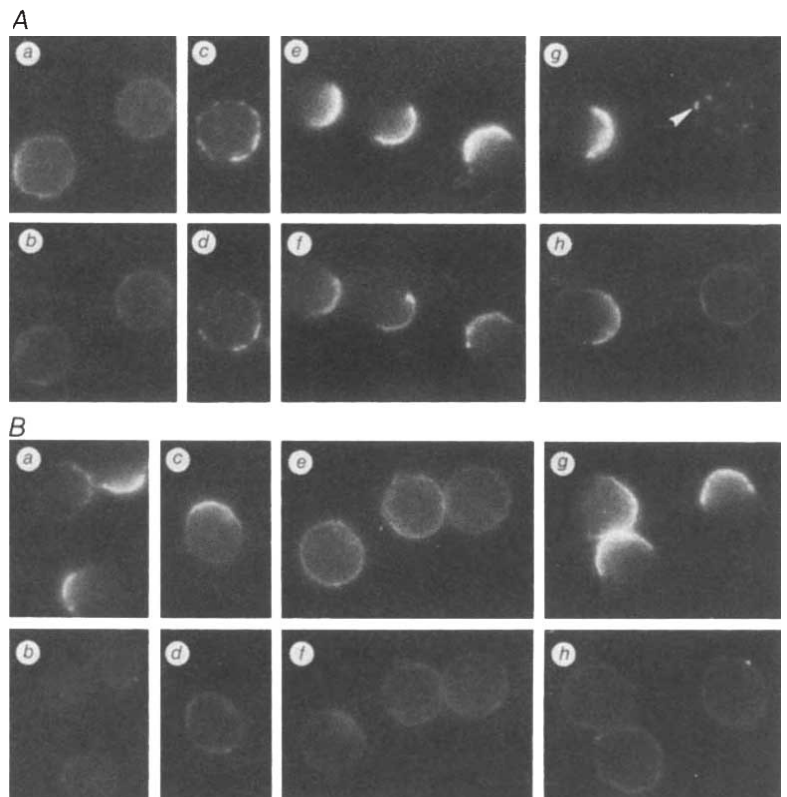
immunoglobulin (Fig. 2A). During the first 5 min of stimulation, the surface immunoglobulin formed aggregates of variable sizes (Fig. 2A, c), and at this stage p21^{ras} was present in a correspondingly patchy distribution (Fig. 2A, d). Within 15 min, the fluorescence staining for p21^{ras} was concentrated in one pole of the cell (Fig. 2A, f) directly under the capped surface immunoglobulin (Fig. 2A, e). In some cells, p21^{ras} was not exclusively localized under the cap but could be detected at much lower intensities under the non-capped regions of the membrane as well (not shown). These results demonstrate that the pattern of p21^{ras} redistribution corresponds to that of surface immunoglobulin at each stage of the capping process. Further incubation of the capped B cells resulted in the disappearance of surface immunoglobulin staining and the appearance of intracellular staining corresponding to internalized GAMIg-surface immunoglobulin complexes (Fig. 2A, g). This is in agreement with earlier studies demonstrating that cap formation is followed by the internalization of ligand-surface-immunoglobulin complexes⁵. The p21^{ras} staining under these conditions, however, was not associated with the internalized immunoglobulin but was uniformly distributed along the cell membrane (Fig. 2A, h). Thus, the coincident movements of p21^{ras} and surface immunoglobulin seem to be confined to the phase of cap formation.

It has been demonstrated that intact (IgG) anti-mouse immunoglobulin antibodies can induce the co-crosslinking of surface immunoglobulin and Fcγ receptors on B lymphocytes

and that the engagement of the Fcγ receptors under these conditions is antagonistic to the surface immunoglobulin-mediated stimulation of B cells^{9,10}. To establish whether the redistribution of p21^{ras} observed with the intact GAMIg reflects the co-crosslinking of surface immunoglobulin and Fcγ receptors, we have examined the effect of F(ab')₂ fragments of GAMIg on p21^{ras} distribution. When added to splenic lymphocytes, fluorescein-conjugated F(ab')₂ fragments of goat anti-mouse immunoglobulin (whole molecule) antibodies (100 µg ml⁻¹) were as effective as the intact antibodies in inducing the redistribution of both surface immunoglobulin and p21^{ras} (not shown). Thus, the co-capping of p21^{ras} and surface immunoglobulin is independent of Fcγ receptors.

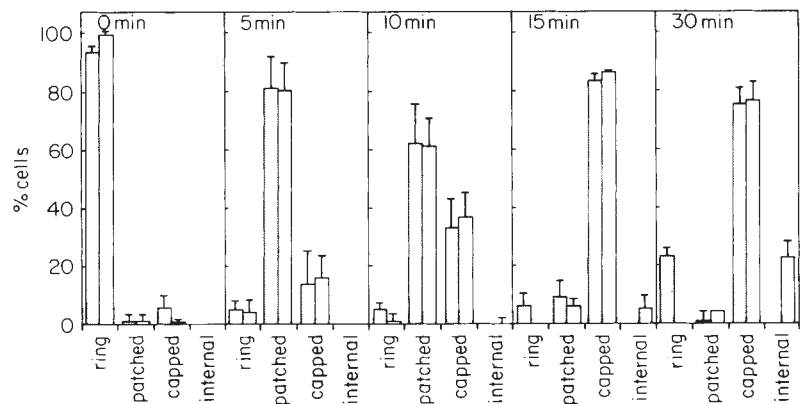
Several experiments were carried out to establish the specificity of the p21^{ras} redistribution during surface immunoglobulin capping (Fig. 2B). First, to ascertain the specificity of the staining *per se*, the primary anti-p21^{ras} antibodies were omitted from the staining procedure (Fig. 2B, a and b). Under these conditions, no significant rhodamine fluorescence was detected (Fig. 2B, b), indicating that the secondary anti-rabbit IgG antibody does not crossreact with mouse B-cell surface immunoglobulin. Second, to examine whether the redistribution of p21^{ras} is a general phenomenon that accompanies the capping of surface molecules, we have examined the staining pattern of p21^{ras} during the capping of concanavalin A (con A) receptors (Fig. 2B, c and d). Treatment of mouse splenic lymphocytes with con A (30 µg ml⁻¹) for 30 min at 37 °C

FIG. 2 A, Redistribution of p21^{ras} during the capping of surface immunoglobulin in mouse splenic B lymphocytes. Upper panels (a, c, e, g) show the fluorescence staining pattern of surface immunoglobulin, and lower panels (b, d, f, h) show the fluorescence staining pattern of p21^{ras}. About 50% of splenic lymphocytes displayed immunoglobulin staining. Among the cells that did not show immunoglobulin staining (that is, non-B cells), a high proportion (~90%) were stained for p21^{ras}. a, b, Uniform distribution before stimulation with fluorescein-conjugated goat anti-mouse immunoglobulin antibodies (GAMIg). c, d, Patching after 5 min of stimulation. e, f, Capping after 15 min of stimulation. g, Internalization of surface immunoglobulin after 30 min of stimulation. Arrow points to intracellular staining corresponding to GAMIg-surface immunoglobulin complexes. h, p21^{ras} staining in the same cells. Note that in the cell displaying internalized immunoglobulin, p21^{ras} is uniformly distributed along the cell surface. The other cell visible in the field displays the co-capping of p21^{ras} and surface immunoglobulin. Identical staining patterns were observed in B-cell-enriched preparations obtained by cytotoxic killing of T lymphocytes using a mouse monoclonal anti-Thy-1.2 antibody and guinea-pig complement followed by fractionation on a 40% Percoll gradient. B, Characterization of the p21^{ras} redistribution. a, b, Specificity control for the indirect immunofluorescence staining. a, Capped surface immunoglobulin at 15 min after stimulation with GAMIg. b, Same cell as in a, stained with secondary antibody alone. c, d, Effect of con A receptor capping on p21^{ras} redistribution. c, Capped con A receptors after 30 min of stimulation with fluorescein-conjugated con A. d, Same cell as in c, stained with anti-p21^{ras} antibodies. e-h, Distribution of the G_i α subunit during surface immunoglobulin capping. e, g, Surface immunoglobulin distribution before (e) and 15 min after (g) stimulation with GAMIg. f, h, Same cells as in e and g, respectively, stained with anti-G_i α subunit antibodies. These antibodies recognize neutrophil G_i, liver G_i and brain G_o α subunits¹¹. **METHODS.** Splenic lymphocytes were collected from BALB/c mice. The cells were suspended in RPMI 1640 containing 2% FCS and plated on polylysine-coated glass coverslips. Induction of patching and capping was carried out at 20 °C using 100 µg ml⁻¹ of fluorescein-conjugated affinity-purified goat anti-mouse IgG (whole molecule) antibodies (Cappel Research Reagents). The induction of capping of con A receptors was carried out using fluorescein-conjugated con A (30 µg ml⁻¹). As con A-induced capping is significantly slower than anti-Ig-induced capping, incubations were at 37 °C. At the noted times, cells were fixed with 1% paraformaldehyde in PBS, pH 7.4 and permeabilized by brief extraction with PBS containing 0.1% Triton X-100.



The coverslips were then incubated with rabbit anti-p21^{ras} antibodies (100 µg ml⁻¹) or with rabbit anti-G_i α subunit antibodies (R16, 17; 20 µg ml⁻¹) for 45 min at 37 °C in PBS containing 10 mg ml⁻¹ BSA, washed for 5 min in PBS and incubated with rhodamine-conjugated goat anti-rabbit antibody (20 µg ml⁻¹) for 45 min at 37 °C in PBS containing 10 mg ml⁻¹ BSA. The coverslips were examined on a Zeiss Axiophot microscope equipped for epifluorescence using a ×63 oil immersion lens. Appropriate excitation and barrier filters were employed to observe the fluorescein or rhodamine fluorescence. Micrographs were taken on Tri-X (Kodak) film and developed in diafine (×4,600).

FIG. 3 Comparison of the time courses of redistribution of surface immunoglobulin and p21^{ras}. Splenic lymphocytes were plated on polylysine coated glass coverslips and incubated with GAMlg (100 µg ml⁻¹) at 20 °C. At the indicated times, the incubation was stopped by the addition of 1% paraformaldehyde. The cells were fixed for 30 min, washed and then further processed for immunofluorescent staining of p21^{ras} as described in the legend to Fig. 1. Cells were first examined for rhodamine (p21^{ras}) staining and then for fluorescein staining or vice versa, and the percentage of lymphocytes showing 'ring', 'patched', 'capped' and internal staining was determined. Staining was scored as 'capped' if the fluorescence was localized at one pole and occupied one-half or less of the surface. Staining was scored as 'ring' when fluorescence was uniformly distributed along the cell periphery and as 'patched' when fluorescence was broken up into patches without obvious polar distribution. Cells were scored for internal staining when fluorescence was found within the cell. Open columns are for p21^{ras} (rhodamine) staining, and shaded columns are for surface



immunoglobulin (fluorescein) staining. At least 300 cells were scored from two independent experiments. The values shown are the mean $\pm$ s.d.

induced the capping of con A receptors (Fig. 2*B, c*) but did not affect the distribution of p21^{ras} (Fig. 2*B, d*), indicating that the redistribution of p21^{ras} is specifically associated with surface immunoglobulin capping. Last, we have monitored the distribution of another signal transducing protein, the G_i α subunit, during surface immunoglobulin capping (Fig. 2*B, e-h*). For these experiments, affinity-purified preparations of antibodies that specifically recognize the α subunit of G_i (R16, 17; ref. 11) were employed. In resting B cells, a uniform peripheral staining of the G_i α subunit was observed (Fig. 2*B, f*). The staining was similar in intensity to that of p21^{ras} and was completely abolished by blocking the antibody with the antigen peptide (not shown). In contrast to the co-capping of p21^{ras} with surface immunoglobulin, the distribution of the G_i α subunit remained unchanged throughout the process of surface immunoglobulin capping (Fig. 2*B, g and h*). These results support the idea that the capping of surface immunoglobulins selectively induces the movement of p21^{ras}.

To define better the relationship between surface immunoglobulin cap formation and p21^{ras} redistribution, we compared the time courses of redistribution of surface immunoglobulin and p21^{ras} following stimulation with GAMlg. As shown in Fig. 3, the kinetics of immunoglobulin and p21^{ras} redistributions were essentially identical, indicating that the movement of p21^{ras} and surface immunoglobulin are coordinated during the process of ligand-induced immunoglobulin capping. In contrast, at a later stage of this process, the movements of surface immunoglobulin and p21^{ras} became uncoupled (Fig. 3; 30 min). Whereas GAMlg-surface immunoglobulin complexes were cleared from the cell surface as a result of internalization, p21^{ras} maintained its location in the cell membrane and redistributed back to the resting uniform pattern. As a result, p21^{ras} might once again be available for subsequent patching and capping events in the same cell.

Capping of surface immunoglobulin is an energy-dependent process^{12,13} that is mediated by the contractile activity of the actomyosin peripheral cytoskeleton¹⁴⁻¹⁶. Consequently, treatment of splenic lymphocytes with metabolic inhibitors or with drugs that impair cytoskeletal function results in the inhibition of surface immunoglobulin capping^{17,18}. Treatment of B lymphocytes with sodium azide (15 mM) or with cytochalasin D and colchicine—drugs that perturb microfilament and microtubule structure^{19,20}, respectively—inhibited the capping of both surface immunoglobulin and p21^{ras} (Table 1). These observations suggest that the anti-immunoglobulin-induced redistributions of surface immunoglobulin and p21^{ras} are regulated by common mechanisms.

Because of their oncogenic potential, p21^{ras} proteins have been studied primarily in the context of mitogenic signalling

and growth control in fibroblasts. The strategy used in most of these studies involved *ras*-transfected cell lines that overexpress various p21^{ras} species. The present study examines for the first time the function of p21^{ras} in B lymphocytes. Furthermore, the experimental approach relies on endogenous p21^{ras}, thus offering a direct insight into the physiological role of p21^{ras}. Our observation that the redistribution of p21^{ras} exactly paralleled the patching and capping of surface immunoglobulin is consistent with the idea that p21^{ras} can become associated spatially with immunoglobulin molecules. A similar spatial relationship has been observed between the redistribution of surface immunoglobulin and some of the cytoskeletal elements that comprise the actomyosin contractile system, for example, actin, myosin, and α -actinin^{14,15,21}. On the basis of biochemical and ultrastructural analyses, it is now largely accepted that the redistribution of these cytoskeletal proteins reflects their function in the mechanism responsible for cap formation²². By analogy, the redistribution of p21^{ras} during surface immunoglobulin capping may indicate that p21^{ras} is a component of the cellular machinery responsible for the ligand-induced capping process. Alternatively, the movement of p21^{ras} may reflect its participation as a transducing protein in the signalling pathway triggered by surface immunoglobulin capping. In this regard, it is important to note that signalling through surface immunoglobulins in murine B cells has recently been shown to be mediated by an as-yet unidentified G protein that is insensitive to pertussis toxin²³⁻²⁵. Further studies of the interaction of p21^{ras} with surface immunoglobulin and with the subcortical actin filament network will contribute to an understanding of the role of p21^{ras} in B-cell activation. □

Received 9 April; accepted 9 July 1990.

- McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. *Nature* **310**, 644-649 (1984).
- Sweet, R. W. *et al.* *Nature* **311**, 273-275 (1984).
- Gibbs, J. B., Sigal, I. S., Poe, M. & Scolnick, E. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5704-5708 (1984).
- Barbacid, M. A. *Rev. Biochem.* **56**, 779-827 (1987).
- Taylor, R. B., Duffus, P. H., Raff, M. C. & dePetris, S. *Nature new Biol.* **233**, 225-229 (1971).
- Unanue, E. R., Perkins, W. D. & Karnovsky, M. J. *J. exp. Med.* **136**, 885-897 (1972).
- Gross, M. *et al.* *Molec. cell Biol.* **5**, 1015-1024 (1985).
- Willingham, M. C., Pastan, I., Shih, T. C. & Scolnick, E. M. *Cell* **19**, 1005-1014 (1980).
- Scribner, D. J., Weiner, H. L. & Moorhead, J. W. *J. Immun.* **121**, 377-383 (1978).
- Philips, N. E. & Parker, D. C. *J. Immun.* **132**, 627-632 (1984).
- Bokoch, G. M., Bickford, K. & Bohl, B. P. *J. Cell Biol.* **106**, 1927-1936 (1988).
- Unanue, E. R., Karnovsky, M. J. & Engers, H. D. *J. exp. Med.* **137**, 675-689 (1973).
- Loor, F., Forni, L. & Pernis, B. *Eur. J. Immun.* **2**, 203-212 (1972).
- Schreiner, G. F., Fujiwara, K., Pollard, T. D. & Unanue, E. R. *J. exp. Med.* **145**, 1393-1398 (1977).
- Gabbiani, G., Chaponnier, C., Zumbé, A. & Bassali, P. *Nature* **269**, 697-698 (1977).
- Bourguignon, L. Y. W., Tokuyasu, K. T. & Singer, S. J. *J. cell Physiol.* **95**, 239-258 (1978).
- Edelman, G. M., Yahara, I. & Wang, J. L. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1442-1446 (1973).
- dePetris, S. *Nature* **250**, 54-56 (1974).
- Wessels, N. K. *et al.* *Science* **171**, 135-143 (1971).
- Borisy, G. G. & Taylor, E. W. *J. cell Biol.* **34**, 525-534 (1967).
- Geiger, B. & Singer, S. J. *Cell* **16**, 213-222 (1979).
- Weihung, R. R. in *Methods and Achievements in Experimental Biology* Vol. 8 (ed. Gabbiani, G.) 42-109 (Karger, New York, 1979).

23. Gold, M. R., Jakway, J. P. & DeFranco, A. L. *J. Immun.* **139**, 3604–3613 (1987).
 24. Harnett, M. M. & Klaus, G. G. B. *J. Immun.* **140**, 3135–3139 (1988).
 25. Monroe, J. & Haldar, S. *Biochim. biophys. Acta* **1013**, 273–278 (1989).

ACKNOWLEDGEMENTS. We thank G. Bokoch (Scripps Clinic, La Jolla, California) for his gift of anti-G α subunit antibody (R16,17) and the R17 peptide. This work was supported by the NIH and the American Cancer Society.

Hypervariable ultra-long telomeres in mice

David Kipling & Howard J. Cooke*

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

TELOMERE structure and behaviour is less well understood in vertebrates than it is in ciliates and yeasts (reviewed in ref. 1). Like all other eukaryotic chromosomes, those of vertebrates terminate in an array of a short repeated sequence. In vertebrates this sequence is (TTAGGG) $_n$, as shown by *in situ* hybridization^{2,3}. In humans, these terminal repeats are heterogeneous in length, averaging about 10 kilobases in blood cells^{4–6}. Here we report the structure and inheritance of the terminal repeats present at mouse telomeres. The (TTAGGG) $_n$ tracts are many times larger than those present at human telomeres. Because of their constancy in length through somatic cell divisions, they are resolved as multiple discrete restriction fragments of up to 150 kilobases. Strikingly, this banding pattern is highly polymorphic within populations of inbred mice, suggesting an unusually high mutation rate. Indeed, although the banding pattern is inherited in a largely mendelian fashion, (TTAGGG) $_n$ tracts of new size appear frequently in family studies.

When human DNA from peripheral blood cells is digested with a frequently cutting restriction enzyme, hybridization with a (TTAGGG) $_4$ oligonucleotide probe detects a range of terminal fragment sizes, averaging around 10 kilobases (kb) (refs 4–6). These are seen as a smear of hybridization signal when separated by conventional agarose gel electrophoresis. By contrast, the hybridization signal with similarly digested mouse spleen DNA is much stronger, and corresponds largely to high relative molecular mass (M_r) DNA at the position of limit mobility (~ 25 kb and above). We estimate from quantitative hybridization (data not shown) that the mouse genome contains roughly 8–16 times more (TTAGGG) $_n$ than the human genome.

To further analyse mouse (TTAGGG) $_n$ tracts, we prepared high- M_r DNA, embedded in agarose plugs, from livers of individual mice. After digestion with a restriction enzyme with a 4-base-pair (bp) recognition sequence, such as *Mbo*I, fragments were separated by pulsed-field gel electrophoresis. What results is not the featureless smear of (TTAGGG) $_n$ hybridization seen with human DNA, but a signal resolving into several discrete bands, the pattern type being strain-dependent (Fig. 1). One pattern type is seen with the C57BL/6J inbred strain. The bulk of the hybridization is to fragments between 20 and 65 kb, and although not resolved into discrete bands, the hybridization signal is suggestive of a cluster of incompletely resolved fragments. Another pattern type is exemplified by the DBA/2 inbred strain, and is similar to that seen with CBA/Ca and BALB/c inbred strains. As well as what may be a cluster of fragments not fully resolved in the 20–70 kb range, there are a small number of often discretely resolved bands between 70 and 140 kb.

These large (TTAGGG) $_n$ -containing fragments are largely devoid of recognition sites for all restriction enzymes tested so far. Large fragments, similar in size, are seen after digestion with *Bam*HI, *Bgl*II, *Eco*RI, *Hind*III, *Hinf*I, *Msp*I, *Hae*III, *Rsa*I, *Mbo*I, *Sau*3A or *Alu*I. (Small differences are seen, some

of which may reflect varying amounts of flanking DNA included in the fragments.) The banding pattern seen with DNA from a CBA/Ca individual digested with five different enzymes, each with a 4-bp recognition sequence, is illustrated in Fig. 2.

The lack of restriction sites within these fragments, even for enzymes with a 4-bp recognition sequence, and our estimate of the (TTAGGG) $_n$ content of the mouse genome, suggest that these large fragments are composed almost entirely of TTAGGG repeats, although we cannot exclude the possibility that they are interspersed with other repetitive sequences lacking restriction sites.

A telomeric location of (TTAGGG) $_n$ in the mouse is indicated by *in situ* hybridization³. In addition, telomeric sequences in intact DNA are preferentially removed by exonuclease *Bal*31 (refs 4–6), and indeed *Bal*31 digestion of intact mouse DNA embedded in agarose plugs does lead to preferential loss of (TTAGGG) $_n$ -containing fragments (Fig. 3). During *Bal*31 digestion, average band size decreases from ~ 80 kb to 20 kb, con-

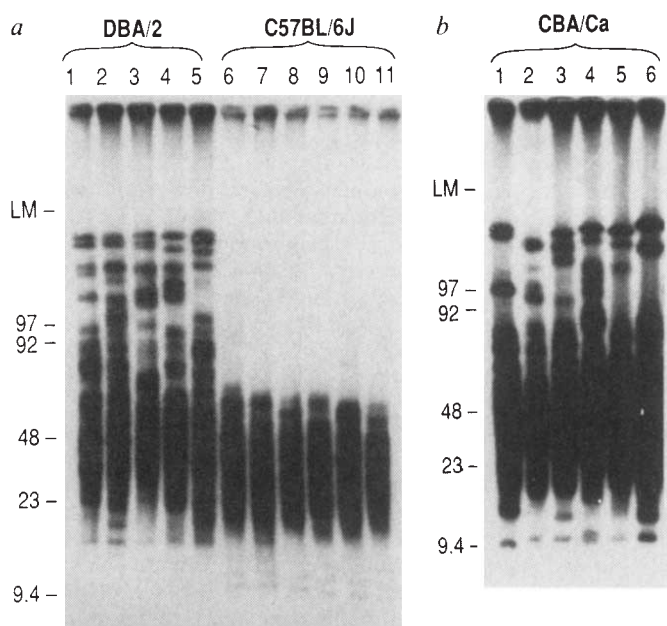


FIG. 1 (TTAGGG) $_n$ tract size polymorphism within inbred mouse strains. DNA from random individuals from colonies of DBA/2, C57BL/6J (a) and CBA/Ca (b) mice was digested with *Mbo*I, separated by pulsed-field gel electrophoresis, transferred to Hybond-N, and probed with ³²P-labelled (TTAGGG) $_4$. All are male except CBA/Ca animal 6.

METHODS. Each mouse liver was homogenized using a dounce homogenizer in 10 ml of 1 $\times$ buffer A (60 mM KCl, 15 mM NaCl, 0.5 mM spermine, 0.15 mM spermidine, 15 mM Tris-HCl buffer, pH 7.5, 14 mM β -mercaptoethanol), 2 mM EDTA, 0.3 M sucrose. The homogenate was passed through three layers of gauze and layered over 15 ml of 1 $\times$ buffer A, 1.37 M sucrose, 1 mM EDTA. After centrifugation in a swing-out rotor for 15 min at 12,000 r.p.m. and 4 $^{\circ}$ C, the pellet of nuclei was resuspended in 2 ml of 1 $\times$ buffer A, 1 mM EDTA. An equal volume of 1% low-gelling-temperature agarose in 1 $\times$ buffer A, 1 mM EDTA (maintained molten at 50 $^{\circ}$ C) was added, and aliquots transferred to plug moulds. When set, the plugs were removed and washed three times by incubation in 0.5 M EDTA, 10 mM Tris-HCl, pH 9.5, 1% *N*-lauroylsarcosine, 0.2 mg ml⁻¹ proteinase K at 50 $^{\circ}$ C for 48 h. Plugs were then stored in 0.5 M EDTA, 10 mM Tris-HCl pH 9.5, 1% *N*-lauroylsarcosine at 4 $^{\circ}$ C. After digestion with *Mbo*I as described¹¹, DNAs were separated on a 1% agarose gel using a Pharmacia LKB Pulsaphor apparatus and a hexagonal electrode. Electrophoresis was in 0.5 $\times$ TBE (0.045 M Tris base, 0.045 M boric acid, 0.001 M EDTA) at 200 V for 21 h with a 5-s switching interval, the running buffer being maintained at 12 $^{\circ}$ C. The gel was transferred to Hybond-N and hybridized to ³²P-labelled (TTAGGG) $_4$ as described⁵. Size markers (kb) derived from concatamers of phage lambda DNA, phage lambda DNA digested with *Hind*III, and *Saccharomyces cerevisiae* chromosomes (strain YPH148; from P. Hieter) are indicated. LM: limit mobility.

* To whom correspondence should be addressed.

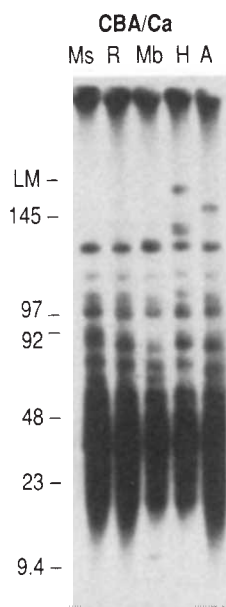


FIG. 2 The banding pattern as visualized with different frequently cutting restriction enzymes. Liver DNA from a male CBA/Ca was digested with *MspI* (Ms), *RsaI* (R), *MboI* (Mb), *HaeIII* (H) or *AluI* (A), separated by pulsed-field gel electrophoresis, and probed with (TTAGGG)₄ as in Fig. 1.

sistent with (TTAGGG)_n, being present along the entire length of these fragments, rather than only at the ends of fragments composed of some other sequence.

The observation that mouse (TTAGGG)_n tracts can be visualized as discrete bands places limits on the extent of length variation during somatic cell divisions. We have no evidence to indicate that individual (TTAGGG)_n tracts in the mouse show length heterogeneity, as is seen in humans⁴⁻⁶. Band widths can be as little as 2 kb; if one assumes a minimum of 25 cell generations to produce a full-sized liver, this suggests that any variation in the length of individual (TTAGGG)_n tracts must be no more than a few hundred base pairs per cell generation, and may be considerably less. A low turnover rate is also suggested by the rate of sequence loss at the tips of *Drosophila melanogaster* chromosomes lacking telomeric sequences^{7,8}.

Random individuals from colonies of C57BL/6J, CBA/Ca and DBA/2 inbred mice are shown in Fig. 1. The banding pattern

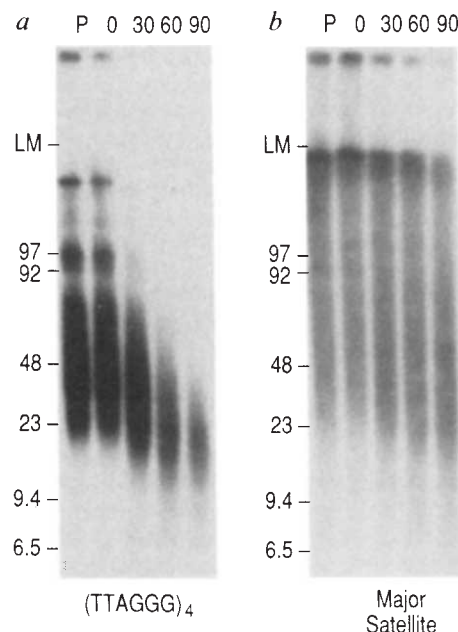


FIG. 3 *Bal31* sensitivity of (TTAGGG)_n tracts in the mouse genome. *a*, DNA embedded in agarose plugs was digested with exonuclease *Bal31* for the times indicated (min) and then separated by pulsed-field gel electrophoresis after digestion with *MspI*. After transfer to Hybond-N, the filter was probed with (TTAGGG)₄. *b*, Filter was stripped and re-probed to detect the mouse major satellite DNA as an internal control.

METHODS. Plugs containing liver DNA from a male DBA/2 were rinsed three times by incubation in 10 mM Tris-HCl, 1 mM EDTA pH 7.5 for 30 min, and once in 1 × *Bal31* buffer (600 mM NaCl, 12 mM MgCl₂, 12 mM CaCl₂, 0.2 mM EDTA, 20 mM Tris-HCl pH 7.5) for 60 min. One pre-treatment ('P') plug was then removed. Remaining plugs were transferred to 1 × *Bal31* buffer containing 28 U ml⁻¹ *Bal31* (BRL) and immediately put on ice, and after 30 min transferred to 37 °C. Plugs were removed after 0, 30, 60 and 90 min and added to 500 μl of 0.5 M EDTA to stop the reaction. Plugs were then rinsed extensively with 10 mM Tris-HCl, 1 mM EDTA pH 7.5 before digestion with *MspI*. Samples were separated by pulsed-field gel electrophoresis and probed with (TTAGGG)₄ as before. The filter was then stripped and re-probed with pMR196, which detects the mouse major satellite¹², as described⁵.

due to the (TTAGGG)_n tracts is strikingly polymorphic between individuals of the same inbred strain. Very few bands are clearly common to all individuals, although the DBA/2 animals share a common 18-kb fragment, and the C57BL/6J animals share

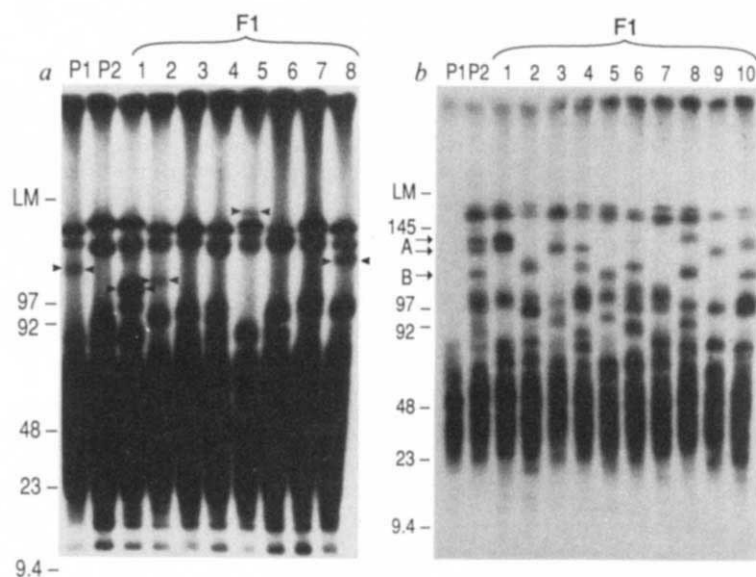


FIG. 4 Family studies illustrating the inheritance pattern of the (TTAGGG)_n tracts and the appearance of new size alleles. *a*, Family consisting of two CBA/Ca animals (P1, male parent; P2, female parent) and a litter from this cross (1-3, male; 4-8, female). All samples are liver DNA digested with *MboI*, separated by pulsed-field gel electrophoresis and probed with (TTAGGG)₄ as in Fig. 1. Arrowheads indicate bands described in the text. *b*, Family consisting of a female C57BL/6J (P1), a male DBA/2 (P2), and a litter from this cross. Parental DNA was obtained from liver, offspring DNA was obtained from whole embryos at 17 days post-coitus. Samples were digested with *MboI*, separated by pulsed-field gel electrophoresis and probed with (TTAGGG)₄ as in Fig. 1. In both *a* and *b* filters were subsequently stripped of the (TTAGGG)₄ probe and re-probed with pMR150, which detects the mouse minor satellite DNA¹². This yields a ladder of about 20 fragments in the 20-150 kb range that acts as an internal control for band-shift artefacts due to partial digestion or aberrant sample migration; no such artefacts were seen (data not shown).

three common fragments between 12 and 16 kb, possibly representing internal genomic fragments.

We have followed the inheritance of these polymorphic bands in several crosses. Figure 4a shows a cross between two CBA/Ca animals; most bands seen in the eight offspring are of a similar size to bands present in the parents. Although we cannot demonstrate an unequivocal relationship between parental and offspring bands, the inheritance pattern is largely consistent with simple mendelian inheritance from both parents. But it is clear from Fig. 4a that there are several bands present in the offspring (indicated by arrowheads) that are of a size dissimilar to that of any parental band; four of the eight offspring possess what seem to be new alleles.

To reduce the number of inherited bands in the 70–140-kb region and thus simplify the analysis, we studied several C57BL/6J × DBA/2 crosses, one of which is illustrated in Fig. 4b. A number of bands, such as two indicated 'A', seem to correspond to one allele of a heterozygous locus, being inherited by roughly half of the offspring. The inheritance of the band marked 'B' is less clear. Six offspring show bands of a similar but not identical size, and although it is a small sample, it seems that the bands are either of two similar sizes. We cannot so far ascertain whether this reflects limited germline mosaicism or selection from a more extensively heterogeneous sperm population. Furthermore, there is a parental band in Fig. 4a, indicated by an arrowhead in parent P1, that is not seen in any of the offspring. This again may reflect mosaicism, further suggested by its apparent presence at a less than unitary amount. But it seems clear that the observed instability of (TTAGGG)_n tract length through generations can account for the high degree of polymorphism seen between individuals of an inbred strain.

It is important to know the site and mechanism of the alteration in length. If variation occurs in somatic tissue, one might be able to detect mosaicism. Our preliminary data from a few animals did not detect any difference in the banding pattern between DNA from liver, brain, spleen or testes of the same animal (data not shown). Mosaicism may be rarely detectable, however, especially as it must occur sufficiently early in a stem-cell lineage so that a new band will be present in a sufficient proportion of cells to allow detection. In addition, highly pure spermatozoan samples obtained by careful dissection show identical banding patterns to liver DNA from the same animal (data not shown). Clearly most of the (TTAGGG)_n tracts remain essentially unchanged on passage through meiosis.

It is not obvious why mice should have such large (TTAGGG)_n tracts. Telomere elongation in the germ line must be closely matched to the rate of sequence loss, if gradual lengthening or shortening over future generations is to be avoided. This implies the existence of a mechanism that can alter the rate of telomere elongation in response to a change in the length of telomeres present in the cell; this could perhaps be controlled by monitoring the total (TTAGGG)_n content of the genome. Whatever the mechanism of this control, it is clear that on an evolutionary timescale it maintains a greater average telomere length in the mouse. This greater average length does not necessarily require a much greater rate of telomere elongation per cell generation in the mouse. Furthermore, a similar length heterogeneity may occur in the mouse as is seen in humans, but is not detected as it is now small in proportion to the much larger telomere size. A genetic control of average telomere length might explain the strain-specific variation in average length illustrated in Fig. 1.

Whether long telomeres are a result of selection or simply a neutral change is not clear. Their size seems largely unchanged on passage to subsequent generations, as well as through somatic cell division, so it is unlikely that the extra length is a defence against rapid loss of sequence. Nor are mouse telomeres significantly reduced in size during the animal's lifespan; a 17-month-old individual still showed normal size distribution of fragments characteristic of its strain (data not shown). This, and

the much longer telomeres of this short-lived species, suggests that telomere shortening is unlikely to have any causal role in ageing *in vivo*, in contrast to some recent speculations⁹. The shortening of human telomeres during ageing *in vivo*¹⁰ may instead indicate that telomere maintenance is another metabolic process that senescent cells are unable to perform as efficiently. □

Received 27 June; accepted 2 August 1990.

1. Zakian, V. A. *Rev. Genet.* **23**, 579–604 (1989).
2. Moyzis, R. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6622–6626 (1988).
3. Meyne, J., Ratliff, R. L. & Moyzis, R. K. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7049–7053 (1989).
4. Allshire, R. C., Dempster, M. & Hastie, N. D. *Nucleic Acids Res.* **17**, 4611–4627 (1989).
5. Cross, S. H., Allshire, R. C., McKay, S. J., McGill, N. I. & Cooke, H. J. *Nature* **338**, 771–774 (1989).
6. de Lange, T. *et al. Molec. cell. Biol.* **10**, 518–527 (1990).
7. Biessmann, H., Carter, S. B. & Mason, J. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1758–1761 (1990).
8. Levis, R. W. *Cell* **58**, 791–801 (1989).
9. Harley, C. B., Futcher, A. B. & Greider, C. W. *Nature* **345**, 458–460 (1990).
10. Hastie, N. D. *et al. Nature* (in the press).
11. Allshire, R. C. *et al. Cell* **50**, 391–403 (1987).
12. Pietras, D. F. *et al. Nucleic Acids Res.* **11**, 6965–6983 (1983).

ACKNOWLEDGEMENTS. We would like to thank Ian Jackson and Nick Hastie for their enthusiastic interest in this project and for many helpful discussions, Arthur Mitchell for DNA preparations of pMR 150 and pMR 196, and Sarah Bowen and Vasker Bhattacharjee for comments on the manuscript. This work was supported by a grant from the Medical Research Council Human Genome Mapping Project.

High specificity of a phosphate transport protein determined by hydrogen bonds

Hartmut Luecke & Florante A. Quiocho*

Howard Hughes Medical Institute, and Departments of Biochemistry and Structural Biology and of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas 77030 USA
Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77251 USA

TRANSPORT of the essential nutrient phosphorus—primarily in the form of orthophosphate—into cells and organelles is highly specific. This is exemplified by the uptake of phosphate or its close analogue arsenate by bacterial cells by way of a high affinity active transport system dependent on a phosphate-binding protein; this system is unable to recognize other inorganic oxyanions and is, moreover, distinct from the one for sulphate transport^{1,2}. The phosphate-binding protein is a member of a family of periplasmic proteins acting as initial high-affinity receptors for the osmotic shock-sensitive active transport systems or permeases for various sugars, amino acids, oligopeptides, and oxyanions^{2,3}. We report here the highly refined 1.7 Å resolution X-ray structure of the liganded form of the phosphate-binding protein. The structure reveals the atomic features responsible for phosphate selectivity, either in monobasic or dibasic form, and the exclusion of sulphate. These features are fundamental to understanding phosphate transport systems and molecular recognition of charged substrates or ions in other biological processes.

Phosphate-binding protein (PBP) consists of a single polypeptide chain (relative molecular mass M_r 34,400 estimated from the sequence of 321 residues)^{4,5} with one tight phosphate-binding site ($K_d = 1 \mu\text{M}$ at pH 8.3)^{6,7}. Whereas arsenate is also transported by the same system¹, sulphate is ineffective as a substrate or inhibitor (binding to PBP at least five orders of magnitude less tightly than phosphate).

Crystallization of PBP at pH 4.5 has already been described⁷. The determination of the PBP structure⁸ is based on initial phasing from single isomorphous replacement from a four-site KI/I₂ derivative with anomalous scattering contribution (Table

* To whom correspondence should be addressed at: Howard Hughes Medical Institute, One Baylor Plaza, Houston, Texas 77030, USA.

TABLE 1 Crystallographic and refinement data

	Native	KI/I ₂
Resolution (Å)	1.7	2.5
Diffraction data*		
Number of observations	102,157	97,758
Number of unique reflections	36,946	11,673
Completeness (%)	95.4	95.9
$\langle I \rangle / \langle \sigma_I \rangle$	22.0	38.6
R_{merge} (%)	5.6	4.2
R_{Bijvoet} (%)		2.1
Phasing statistics		
Number of Bijvoet pairs		7,893
Number of heavy atom sites		4
r.m.s. F_H /residual		2.03
R_{Cullis}		0.57
r.m.s. R_{anom}		0.56
Mean figure of merit		0.69
Restrained least-squares refinement statistics		
R_{cryst} (%)	14.7	
Protein atoms	2,439	
Phosphate atoms	5	
Water molecules	259	
r.m.s. deviations from ideal stereochemistry		
Bond distances (Å)	0.016	
Angle distances (Å)	0.037	
Dihedral distances (Å)	0.045	
Planar group distances (Å)	0.010	
Peptide ω -angle (°)	2.0	
Chiral volume (Å ³)	0.147	

* Each data set was collected from one crystal using an ADSC two area detector system mounted on a Rigaku RU200 rotating anode. The native and derivative crystals are isomorphous: space group $P2_12_12_1$ with cell dimensions $a = 41.97$ Å, $b = 63.98$ Å, $c = 123.97$ Å. The KI/I₂ derivative was obtained following the procedure described in ref. 17. R_{merge} is the conventional R -factor on intensities for symmetry-related reflections. R_{Bijvoet} is the R -factor on intensities for Bijvoet pairs.

$$R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|; \quad R_{\text{Cullis}} = \sum ||F_{\text{PH}} \pm F_{\text{P}}| - F_{\text{H}}| / \sum |F_{\text{PH}} - F_{\text{P}}|;$$

$$\text{r.m.s. } R_{\text{anom}} = [\sum (\Delta F_{\text{obs}}^{\pm} - \Delta F_{\text{calc}}^{\pm})^2 / \sum (\Delta F_{\text{obs}}^{\pm})^2]^{1/2}.$$

1), solvent-flattening⁹, model building, phase combination¹⁰, simulated annealing¹¹, and, finally, restrained least squares refinement¹² at 1.7 Å resolution which yielded an R -factor of 0.147. The crystallographic results are summarized in Table 1 and the legend to Fig. 1. We have also obtained a well-refined

1.8 Å isomorphous structure of PBP at pH 6.2 which is virtually identical to the pH 4.5 structure. The 1.7 Å electron density of the phosphate-filled substrate-binding site is exceptionally well defined (Fig. 1a). This is consistent with analysis, which shows that the central region of the molecule surrounding the binding site has the lowest mean thermal B value (Figs 1b and 2a) and, hence, has the most accurate atomic coordinates.

Binding proteins vary in size (M_r values 25,000–50,000) and in sequence^{2,3}. Nevertheless, the structure of PBP (Fig. 1b) is similar overall to the six other binding protein structures determined in our laboratory^{2,13}. (A detailed description of the structure will be published elsewhere.) It consists of two separate similarly folded globular domains. The deep cleft between the domains contains the bound phosphate which is totally dehydrated and completely sequestered ~8 Å below the protein surface.

The buried phosphate is held tightly in place by 12 strong hydrogen bonds, all formed with protein groups—5 NHs of the main chain, 2 NHs of the Arg 135 side chain (which is also salt-linked with Asp 137), 4 OHs of two serines and two threonines, and 1 oxygen of Asp 56 side chain (Fig. 2 and Table 2). Three of the phosphate oxygens (O1, O2, and O3) are heavily involved by accepting ten hydrogen bond donors, six of which originate from a seven-residue segment folded into a loop (residues 135–138) and the first turn of helix VI (139–141). The O2 is the recipient of four hydrogen bonds in a rare pyramidal arrangement with the heavy atoms (2 N and 2 O) of the four donor groups forming the base and the O2 atom the apex. There is pairing of all polar groups and total involvement of all donatable protons in the binding site region, giving rise to extensive networks or arrays of hydrogen bonds, and to precisely and stably oriented functional residues.

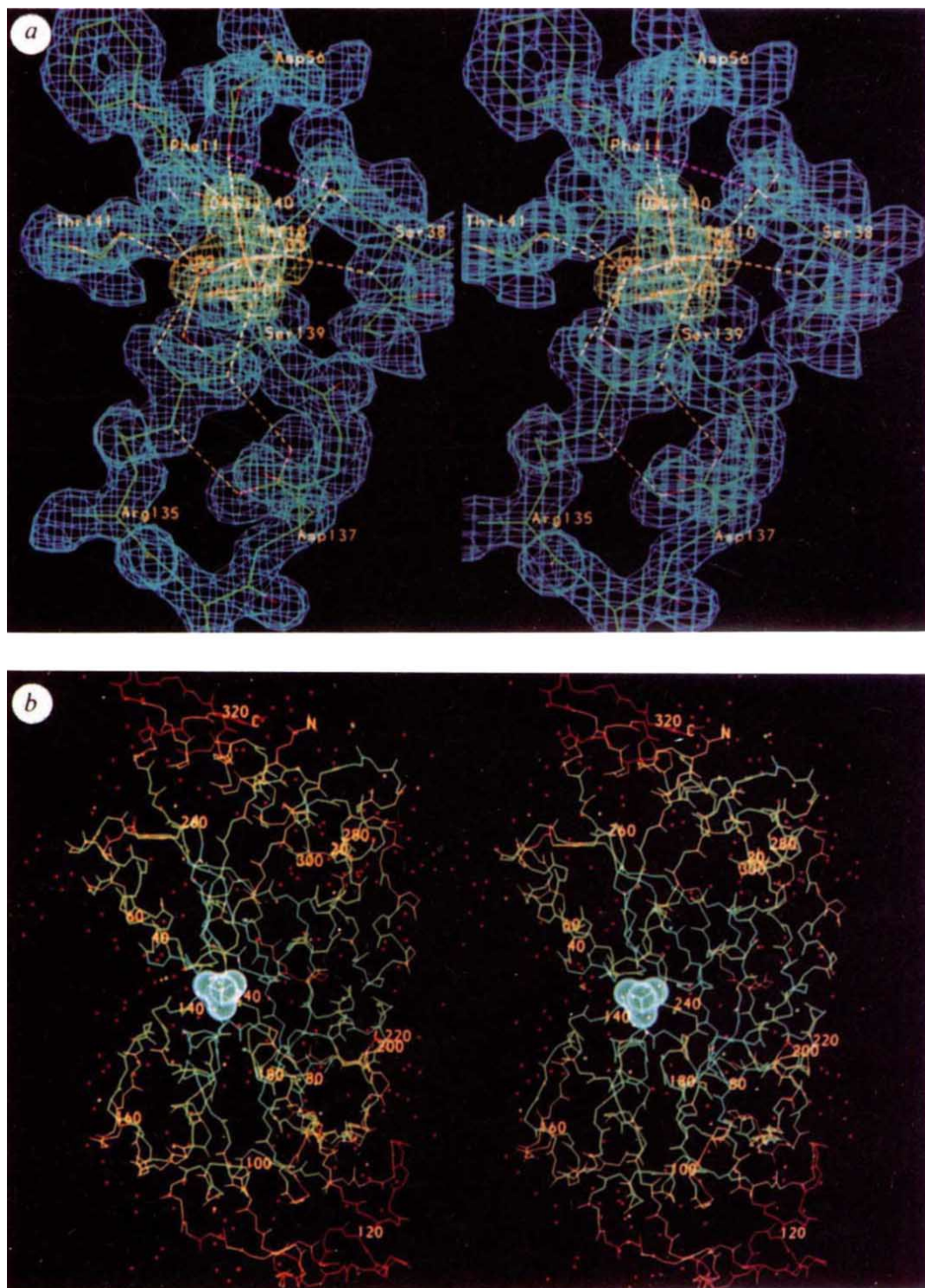
The interactions associated with the phosphate O4 are not only distinctive but are also the key to molecular understanding of the exquisite specificity of PBP (Fig. 2). The O4 accepts a hydrogen bond from an NH of a backbone peptide unit whose carbonyl oxygen is in turn the recipient of two hydrogen bonds. It is also in close contact (2.45 Å) with O_{δ2} of the Asp 56 carboxylate side chain. This close association demands that the two oxygens share a proton on the O4 phosphate, resulting in a hydrogen bond between the phosphate and the carboxylate side chain. This accounts for the slight preference for dibasic phosphate^{6,7}. Also, the hydrogen bonding between the phosphate and carboxylate is reminiscent of the short H bonds in the carboxylate-carboxyl interactions found in small compounds¹⁴ and in proteins¹⁵.

TABLE 2 Hydrogen bonds between phosphate and PBP

Phosphate atom	Domain	Hydrogen bond partner		Distance (Å)		Angle (°)
		Residue	Group	H...Y	X...Y	X—H...Y
O1	N	Thr 10	NH	1.85	2.81	156.0
	N	Thr 10	O _γ H	1.73	2.69	167.2
	C	Arg 135	N _{η1} H	1.82	2.84	172.3
O2	C	Arg 135	N _{η2} H	1.96	2.90	148.9
	C	Ser 139	O _γ H	1.80	2.77	169.0
	C	Thr 141	NH	1.84	2.83	159.2
	C	Thr 141	O _γ H	1.79	2.68	149.7
O3	N	Ser 38	NH	1.68	2.67	163.9
	N	Ser 38	O _γ H	1.69	2.63	160.9
	C	Gly 140	NH	1.74	2.76	177.1
O4	N	Phe 11	NH	1.92	2.92	163.3
O4—H	N	Asp 56	O _{δ2}	1.57	2.45	151.2
Mean				1.78 (0.11)	2.75 (0.13)	161.6 (8.7)

In order to obtain hydrogen bond distances and angles, all hydrogen atoms were placed and energy-minimized with the program AMBER 3.0²³ while freezing the refined positions of the heavy atoms and heavily constraining the hydrogens fixed by geometry (such as those in peptide units, Arg, etc). X is the hydrogen bond donor group and Y the acceptor group (see Fig. 2b). Note that the hydrogen bond between the phosphate O4—H and Asp 56 O_{δ2} is significantly shorter than the mean value. Also, the main chain NH and side chain OH groups of three of the hydroxyl-containing residues (10, 38, 141) simultaneously serve as H-bond donors, and the hydroxyl oxygen (X) to phosphate oxygen (Y) distances tend to be short.

FIG. 1 *a*, Stereoscopic view of the binding site of PBP with bound phosphate. This view shows the 1.7 Å electron density (calculated with $(2|F_o| - |F_c|)$ coefficients and final α_c phases) and superimposed refined model. The density surface (phosphate in yellow and protein in blue) was contoured at $0.5 \text{ e}/\text{\AA}^3$ and a grid-spacing of $0.5 \text{ \AA}$. Dashed lines represent hydrogen bonds. CHAIN, a density-fitting and molecular graphics program developed by Dr J. S. Sack, was used in this study. *b*, The refined main chain atoms of PBP (total of 321 residues), the phosphate molecule shown with van der Waals surface and water molecules represented by dots. Every 20th α -carbon is numbered, and N and C designate the amino- and carboxy-terminal ends, respectively. The atoms are coloured according to their temperature factors: blue for B -values below $10 \text{ \AA}^2$, green between $10 \text{ \AA}^2$ and $15 \text{ \AA}^2$, yellow between $15 \text{ \AA}^2$ and $20 \text{ \AA}^2$, and red over $20 \text{ \AA}^2$. Starting from the binding site region which has the lowest average B -factor (Fig. 2*a*), the B -factor increases radially so that eventually both extremities of the ellipsoidal protein exhibit the highest motion. The molecule has dimensions of $35 \text{ \AA} \times 40 \text{ \AA} \times 70 \text{ \AA}$, and consists of two domains (designated N and C) which are connected by two different peptide segments. The first domain (N domain) is made up of two segments (residues 1–72 and 257–321) and the second domain (C domain) is composed of a single segment (90–209).



The structure also reveals a simple molecular mechanism for binding—depending on the pH—dibasic (HPO_4^{2-}) or monobasic (H_2PO_4^-) phosphate (see Fig. 2*a, b*). In binding the monobasic species at acidic pH, Ser 38 γ -OH is the only other group favourably positioned to accept—through its oxygen lone pairs—the second phosphate proton (on O3) whilst donating its proton to Asp 56 O₈₂ (distance $3.28 \text{ \AA}$) (Fig. 2*a*). In this binding mode, the negative charge on the phosphate will be concentrated on the two oxygens (O1 and O2) interacting with Arg 135. At pH around 8, where slightly tighter binding for the dibasic species is observed, the hydrogen-bonding interactions involving the Ser 38 γ -OH switches to that shown in Fig. 2*b*. In these interactions, the γ -OH forms a bifurcated hydrogen

bond by also donating to the unprotonated phosphate O3 (distance $2.62 \text{ \AA}$).

Asp 56 has at least three essential roles. It is responsible for the recognition of a proton on the phosphate. By accepting a hydrogen bond from the Ser 38 hydroxyl, Asp 56 facilitates the hydroxyl oxygen in accepting a second phosphate proton. Lastly and of critical importance, Asp 56 is the only residue capable of totally discriminating sulphate or similar tetrahedral fully ionized divalent oxyanions. Although these oxyanions have structural features very similar to those of phosphate¹⁶, they would be repulsed by the negatively charged side chain.

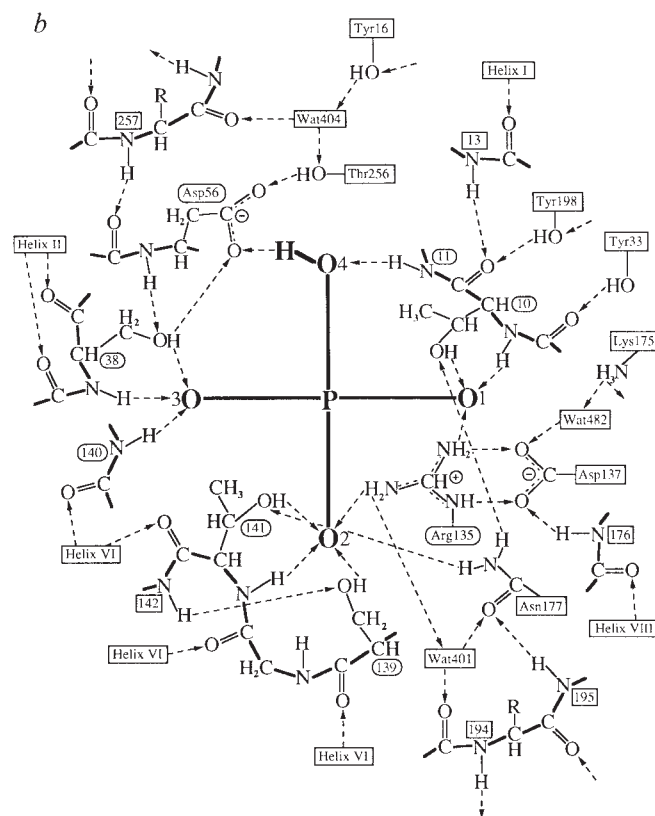
Typical of phosphate transport systems, sulphate or similar fully ionized oxyanions cannot substitute for phosphate or

FIG. 2 Hydrogen-bonding interactions in the PBP-phosphate complex. *a*, A stereo view of the hydrogen bonds (dashed lines) between PBP and dibasic phosphate. This figure is identical to Fig. 1*a*, without the electron density contour. There are no metallic cations in the binding site, and the closest water molecule to the phosphate is at a distance of 5.3 Å. The mean isotropic *B*-factor for the atoms shown is 8.3 (2.1) Å² which is to be compared with 19.4 (11.5) Å² for all non-hydrogen protein atoms. The hydrogen bond parameters are listed in Table 2. Note the salt link between Arg 135 and Asp 137. The hydrogen bond between the donor Ser 38 γ-OH and acceptor Asp 56 O_{δ2} is coloured differently (magenta) to indicate its role in binding monobasic and dibasic phosphates (see text). PBP binds phosphate maximally at pH about 8.3, but only about threefold decrease in binding is observed at pH 5 and 9.5

(B. L. Jacobson and F.A.Q., unpublished; ref. 6). This suggests binding of both monobasic (H₂PO₄⁻) and dibasic (HPO₄²⁻) phosphate, although the latter may be slightly preferred. In the binding of dibasic phosphate, the Ser 38 γ-OH forms a bifurcated hydrogen bond with the phosphate O4 and Asp 56 O_{δ2} (3.28 Å distance). In the binding of monobasic phosphate, Ser 38 hydroxyl accepts a hydrogen bond from the protonated O3 and donates to Asp 56. *b*, Schematic diagram of the hydrogen bonds between PBP and dibasic phosphate and the networks of hydrogen bonds involving other groups in the binding site. With the exception of the peptide carbonyl oxygen of residue 9 which is hydrogen bonded to Tyr 33, the other four peptide units directly associated with the phosphate are further coupled to hydrogen-bond arrays involving other main chain peptide units (indicated by arrows). The two left-handed arrays of alternating hydrogen-bonded peptide units associated with helices II (residues 37–47) and VI (139–150) lead to the bulk solvent (Figs 1*b* and 2*b*; see also refs 17–18). The N termini of helices II and VI are near the phosphate.

arsenate. In fact, the uptake of sulphate by the bacterial high-affinity system requires a sulphate-binding protein (SBP)^{2,3,17}. In contrast to PBP, the sulphate-binding protein binds only fully ionized oxyanions, including chromate and selenate^{16,17}. We have previously refined the 2 Å structure of the SBP-sulphate complex^{18–19}. Many features of the atomic interactions between SBP and sulphate are present in the PBP-phosphate complex. The totally desolvated sulphate is completely sequestered in a similar cleft. The main interaction is through hydrogen bonds with the four sulphate oxygens accepting a total of seven hydrogen bonds—one serine OH, five main chain NH groups, and one NH of a Trp side chain. Hydrogen bond arrays are formed, radiating from all four sulphate oxygens. Finally, in the SBP-sulphate complex and in the PBP-phosphate complex, the constellation of hydrogen bond dipoles and other nearby local dipoles (including those from the N termini of helices or other arrays) are responsible for stabilizing or dissipating the charges on the bound oxyanions (see also ref. 20). The isolated negative charge on the buried Asp 56 of PBP is similarly stabilized (Fig. 2*b*).

Despite these many common features, there is one crucial difference between the phosphate and sulphate receptors: whereas a carboxylate side chain—having key roles as indicated above—is present in the binding site of PBP, it is absent in the SBP. In fact, there is no group in the binding site of SBP that



is suited to be a hydrogen bond acceptor¹⁶. Thus substrate recognition critically hinges on the protonated or deprotonated state of the oxyanions and on the presence or absence of hydrogen-bond acceptor groups in the binding site.

These findings provide greater understanding of electrostatic interactions in molecular recognition of charged ligands and in protein structures (related results refs 13, 19–20). Charges on buried groups can be stabilized by means other than salt linkages. Hydrogen bonds, involving primarily main chain peptide and

hydroxyl groups, are of importance in conferring specificity and affinity for negatively or positively charged ligands. These bonds are also desirable for efficient active transport and rapid ion movement^{2,13,19-20}. Although Arg 135 participates in phosphate binding, salt-linking with Asp 137 (Fig. 2) facilitates transport by diminishing the charge-coupling interaction between the guanidinium and phosphate. This novel mechanism is likely also to be important in ion recognition and rapid movement (see below). There is widespread involvement of hydroxyl side chains in hydrogen bonding interactions with negative and positive groups of charged substrates and side chains as well (the buried Asp 56 and phosphate are best examples, Fig. 2; see also ref. 20). A survey of several refined protein structures reveals similar involvement of hydroxyl-containing residues. These observations are related to the bifunctional nature of hydroxyls; mimicking the dipolar properties of water molecules and having rotational degrees of freedom, hydroxyls are able to accept (through the lone pairs) or donate hydrogen bonds or to wholly or partially substitute hydration of anions or cations. Ion channel receptors, especially those which are gated by ligands such as acetylcholine, glycine and GABA, contain hydroxyl-rich regions (reviewed in ref. 21). Also, channel receptors, especially those that are voltage gated (such as sodium, potassium and calcium channels; reviewed in ref. 22), have regions with high concentrations of charged residues. By being salt-linked (like Arg 135 or Asp 137, Fig. 2) or by dissipating isolated charges, the strength of ion binding by the charged residues in these receptors will be moderated to facilitate fast ion movement. The assemblage of polar groups from these key side chains and/or the main chain is also likely to be important in the selectivity and mechan-

isms of action of these channel receptors and in a variety of biological processes as well (see also ref. 20).

In conclusion, our results not only constitute prime molecular evidence for the high phosphate specificity of a transport receptor protein but also demonstrate the power of hydrogen bonds in enabling proteins to distinguish between phosphate and sulphate. These findings provide new insights into charged ligand recognition and rapid rates in active transport, a process essential to living cells. □

Received 22 May; accepted 20 July 1990.

- Medvecsky, N. & Rosenberg, H. *Biochim. biophys. Acta* **241**, 494-506 (1971).
- Quirocho, F. A. *Phil. Trans. R. Soc. B* **341**, 341-351 (1990).
- Furlong, C. E. in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (ed. Neidhardt, F. C.), 768-796 (American Society for Microbiology, Washington, DC, 1987).
- Surin, B. P. et al. *J. Bact.* **157**, 772-778 (1984).
- Magota, K. et al. *J. Bact.* **157**, 909-917 (1984).
- Medvecsky, N. & Rosenberg, H. *Biochim. biophys. Acta* **211**, 158-168 (1970).
- Kubena, B. D., Luecke, H., Rosenberg, H. & Quirocho, F. A. *J. Biol. Chem.* **261**, 7995-7996 (1986).
- Luecke, H. thesis, Rice Univ. (1990).
- Wang, B. C. *Meth. Enzym.* **115**, 90-112 (1985).
- Read, R. *Acta crystallogr.* **A42**, 140-149 (1986).
- Bruenger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458-460 (1987).
- Hendrickson, W. A. *Meth. Enzym.* **115**, 252-270 (1985).
- Sack, J. S., Saper, M. A. & Quirocho, F. A. *J. molec. Biol.* **206**, 171-191 (1989).
- Speakman, J. C. *Structure & Bonding* **12**, 141-199 (1972).
- Sawyer, L. & James, M. N. G. *Nature* **295**, 76-80 (1982).
- Jacobson, B. L. & Quirocho, F. A. *J. molec. Biol.* **204**, 783-787 (1988).
- Pardee, A. B. *J. Biol. Chem.* **241**, 5886-5892 (1966).
- Pflugrath, J. W. & Quirocho, F. A. *Nature* **314**, 257-260 (1985).
- Pflugrath, J. W. & Quirocho, F. A. *J. molec. Biol.* **200**, 163-180 (1988).
- Quirocho, F. A., Sack, J. S., & Vyas, N. K. *Nature* **327**, 561-564 (1987).
- Barnard, E. A., Garlison, M. G. & Seeburg, P. *Trends Neurosci.* **10**, 502-509 (1987).
- Cattrell, W. A. *Science* **242**, 50-61 (1988).
- Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. *J. comp. Chem.* **7**, 230-252 (1986).

ACKNOWLEDGEMENTS. This work was supported by the Howard Hughes Medical Institute and the NIH.

***In vitro* genetic analysis of the *Tetrahymena* self-splicing intron**

Rachel Green, Andrew D. Ellington & Jack W. Szostak

Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

THE availability of methods for the amplification of nucleic acid sequences^{1,2} allows the genetic analysis *in vitro* of the structural and functional properties of many nucleic acids. We have now developed an *in vitro* selection and amplification system, and used it to analyse the self-splicing *Tetrahymena* ribozyme. A much wider range of selective conditions can be used *in vitro* than is possible with standard *in vivo* methods, and many more variants can be handled *in vitro* than *in vivo*. This method can be used to isolate the wild-type ribozyme, and structural variants that are as active as the wild type, from a pool of over 250,000 variants in only three cycles of selection and amplification.

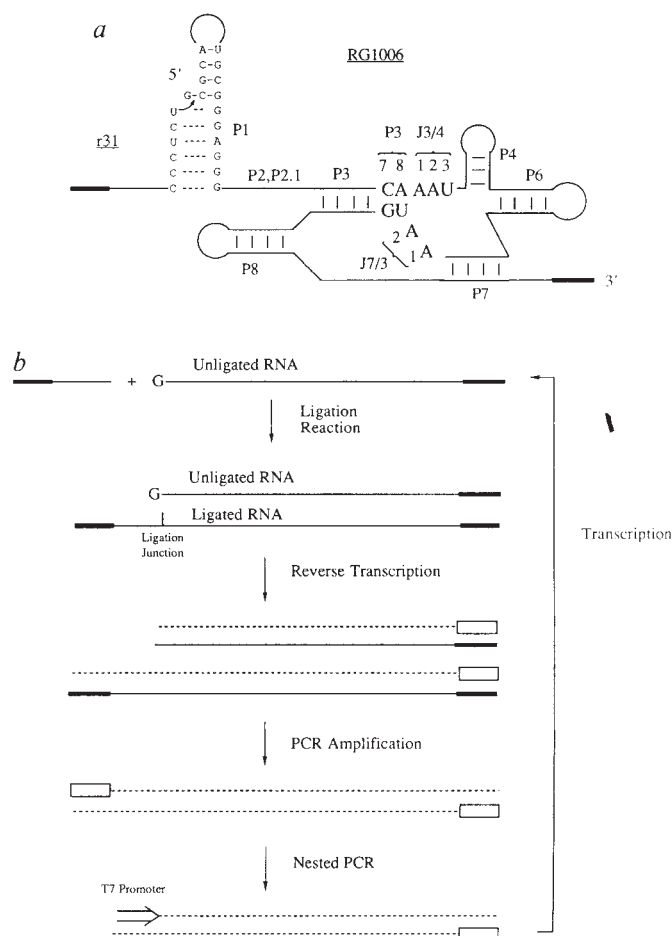
We have developed an iterative procedure for the selective enrichment of catalytically active RNA molecules, by combining *in vitro* RNA replication with an *in vitro* selection system. The phenotype on which the *in vitro* selection is based is an RNA-catalysed RNA ligation reaction. The ribozyme involved is a derivative of the *Tetrahymena* self-splicing intron³ that ligates itself to an RNA oligonucleotide complementary to the internal guide sequence⁴ (Fig. 1a). This reaction is simply the reverse of the first step in self-splicing, with the added oligonucleotide representing the 5' exon.

The added sequence allows selective amplification of catalytically active molecules, as shown in Fig. 1b. The RNA is first copied into complementary DNA by reverse transcriptase extension of a primer complementary to its 3' end. A second primer that hybridizes exclusively to the added sequence information is then used for exponential amplification of the subset of molecules that have undergone the RNA-catalysed ligation reaction. Inactive molecules that fail to ligate cannot be

amplified. DNAs equivalent to the original template are regenerated by a second round of polymerase chain reaction (PCR) with a different 5' primer that carries the T7 promoter and reinserts the 5' guanosine lost during the ligation reaction. RNAs transcribed from this DNA are enriched in catalytically active species; iteration of this procedure through many cycles allows the purification of active enzyme variants, even when these are present at very low abundance in a large pool of inactive variants.

To provide a stringent test of this system, we generated a large pool of sequence variants in which wild-type was present at an extremely low level. This was done by completely randomizing the bases in a region of the ribozyme where phylogenetic comparisons of group I introns had revealed an interesting potential covariation. This region included the 3' end of the base-paired stem P3 and the adjacent joining regions J7/3 and J3/4 (Fig. 1a). In 32 out of 35 group Ib introns, the base pair at P3-7 is C-G, the nucleotide at J7/3-2 is an A, and J3/4 contains three bases; however, in three exceptional introns, P3-7 is G-C, J7/3-2 is U or G, and J3/4 contains only two bases (ref. 5; and F. Michel personal communication). In addition, the prevalence of C-G at P3-7 in the group Ib introns can be contrasted with the variety of base pairs at P3-7 in the group Ia introns. On the basis of these phylogenetic data it seemed plausible that nucleotides in these juxtaposed regions interacted with each other. We then used the *in vitro* selection scheme to establish the rules of variation for this region by selecting for all of its possible functional variants.

A pool of sequence variants for this region was generated by randomizing the nine bases that make up P3-7, P3-8, J7/3 and J3/4. Two primers that spanned these regions were prepared with random (equimolar mix of A, T, C and G) sequences at only the nine chosen bases. These two primers were used for PCR amplification of a DNA fragment corresponding to an internal portion of the ribozyme, using 10 ng of plasmid pRG1006 as a template. After amplification, ~1 µg of the fragment was gel-purified, digested with the restriction enzymes *EcoRI* and *BglII*, and ligated into plasmid cut with *EcoRI* and *BglII* containing the remainder of the ribozyme sequence. The



plasmid used, pRG1007, is a derivative of pRG1006 that contains a stuffer fragment in place of the wild-type *EcoRI*, *BglII* fragment, to avoid contamination of the mutant pool with wild-type sequences; we estimate that carry-over from the wild-type template is less than or equal to the abundance of the wild-type sequence in the mutant pool. Templates suitable for transcription by T7 RNA polymerase were generated directly from the ligation mix by PCR. Sequencing of this pool of DNAs revealed an essentially random mixture of bases at the chosen positions, whereas the remainder of the enzyme was the expected wild-type sequence (Fig. 2a). We estimate that this pool, and the resulting RNA pool, are derived from roughly 10^{10} to 10^{11} molecules of ligated DNA; all of the 4^9 ($= 262,144$) sequence variants should therefore be well represented.

The pool of variants was subjected to three rounds of selection in parallel with a wild-type control. Samples were collected while the rate of ligation was linear with time to maximize

FIG. 1 Scheme for *in vitro* selection and amplification of active ribozymes. **a**, The RNA oligomer r31 interacts with the ribozyme RG1006 by base pairing; the two can become covalently joined by a ribozyme-catalysed ligation reaction involving displacement of the 5' G of the ribozyme as indicated by the arrow. The wild-type sequence of the nine-base region containing P3-7, P3-8, J7/3 and J3/4 is shown here in the context of the conserved elements of group I intron secondary structure⁹. This region has been randomized in the RNAs used for *in vitro* selection. **b**, Functional variants of the ribozyme undergo RNA-catalysed ligation but inactive variants remain unaltered. Reverse transcriptase generates long and short cDNAs from ligated and unligated RNAs; only the long cDNAs are amplified by PCR. A second round of PCR recreates the unligated ribozyme sequence and adds a T7 promoter so that RNA can be generated for subsequent rounds of selection.

METHODS. The RNA oligomer used for ligation (r31, GGGAGCGAAGGC-ACGCCAGCAGGAACCCUCU) was purified on a 15% acrylamide-7 M urea gel⁵ after transcription from a synthetic DNA oligomer containing a T7 promoter sequence¹⁰. The ribozyme RG1006 is a T7 transcript of plasmid pRG1006, which is a derivative of pAG100 (ref. 11) that has been modified to contain P2, P2.1 and part of P1. RG1006 and its mutagenized variants were transcribed from PCR-generated templates¹ and isolated on 4% acrylamide-7 M urea gels. Equimolar ($0.3 \mu\text{M}$) concentrations of RG1006 and r31 were ligated by incubation in $10 \mu\text{l}$ of 10 mM NH_4Cl , 30 mM Tris-HCl buffer, pH 7.4, and either 5 mM or 20 mM MgCl_2 at 45°C for 15 min. RNA (1/100 of the reaction mix) was reverse transcribed using a DNA primer complementary to the 3' end of RG1006 (r31.18, AGCTGGATCCGCTAGCCGACTATATCTTATG, which contains a *Bam*HI site for cloning, or r26.4, which is five bases shorter at its 5' end and was used to generate run-off transcripts). Reverse transcription was carried out with RNase H minus Superscript (BRL) at 65°C for 5 min. Complementary DNA products were amplified by PCR after addition of a second primer (r34.11, GGACTGCAGGGAGCGAAGGCAGCCAGCAGGA). PCR products (362 base pairs) were isolated from 1.5% $1 \times \text{TAE}$ lo-melt agarose gels (FMC Bioproducts, Maine). To regenerate RNAs for subsequent cycles, templates were prepared by nested PCR using r26.4 and an internal primer (r39.18, TAATACGACTCACTATAGCGCAATAGCCTCGAGTACTGC) that recreates the pre-ligation ribozyme sequence and adds the T7 promoter.

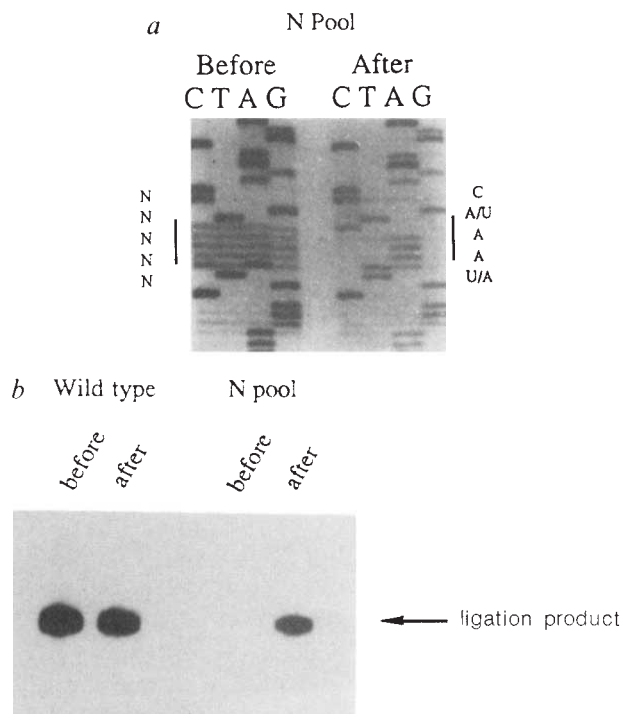


FIG. 2 Enrichment of active ribozyme species through multiple rounds of selection and amplification. **a**, Small segments of sequencing gels showing the P3 and J3/4 regions. The lanes on the left show the original pool of variants, with five randomized bases in this region. To the right, the same region after three rounds of selection, showing the enrichment of a few active species. **b**, Segment of a gel showing RNA-catalysed ligation activity before and after three rounds of selection.

TABLE 1 Mutagenized pools are enriched in active ribozymes by *in vitro* selection

Round	MgCl_2 concentration	Activity enrichment	
		Mutagenized	Wild type
1	20 mM	2.0	1.2
2	20 mM	2.3	1.2
3	5 mM	22.0	1.1
Total		101.2	1.6

Pool activities were measured before and after each round of selection; the ratio of these two values is the activity enrichment for a given round. The activity enrichment is a lower bound for the decrease in sequence complexity of the initial pool. Pools were assayed by determining the rate of ligation of labelled r31 oligonucleotide to RG1006 ribozyme. Assays were carried out as described in Fig. 1, at the indicated MgCl_2 concentration.

selection for the most active variants. The stringency of selection was increased on the third cycle by decreasing the MgCl_2 concentration. After each round of selection, the activity of the pools was tested; the activity of the wild-type control remained essentially constant over the course of the selection, but the mutant pool was enriched in ligation activity following each round (Table 1). After only three cycles of *in vitro* selection and amplification, the mutant pool was 50% as active as the wild-type pool (Fig. 2b). Sequence analysis of the final pool revealed a striking decrease in complexity (Fig. 2a).

After the third round of selection, individual members of the pool were cloned and sequenced (Table 2). The fact that the most common sequence recovered was wild-type suggests that the wild-type sequence is the best of all of the possible 262,144 combinations of the randomized nine bases. Alignment of the remaining sequences provides an 'artificial phylogeny'; comparison of these sequences reveals nonrandom patterns at several positions. The least variation is observed at J3/4-1 and J3/4-2, where only wild-type sequences were recovered; these positions seem to be critical for enzyme activity, and changes at these positions cannot be compensated for by any other changes within the randomized region. In contrast, position J7/3-1 is much more variable. The semi-conserved nature of nucleotides P3-7, P3-8, J3/4-3 and J7/3-2 allows us to test several structural hypotheses.

It is clear that position 7 of stem P3 must be a Watson-Crick base pair, and that C·G is the preferred base pair. Base pairs other than C·G result in a ribozyme that is less active than wild type, as seen before from directed mutagenesis experiments⁵, suggesting that C·G at this position engages in a tertiary interaction with some other base. We do not see any covariation between this base pair and J7/3-2 or any base in the randomized region; our data do not, therefore, provide any support for the putative covariation detected in the phylogenetic data (see above). In contrast, non-Watson-Crick pairings can substitute for the A·U at P3-8. The distribution of bases at this position is nonrandom: the most common pairing is A·U (6), but A·A (4), A·G (3) and U·A (4) are also seen. These data are most

Semi-Conserved Position P3-8

5'	G	A	C	U
3'	G	3		
	A	4		4
	C			
	U	6		

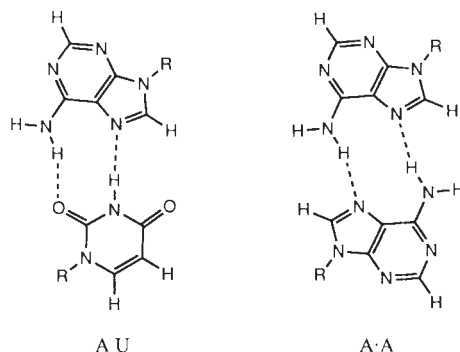


FIG. 3 Structural models of the P8 base pair tested by *in vitro* selection. The matrix (top) shows the nonrandom distribution of potential base pairs at P3-8 for functional ribozyme variants generated by *in vitro* selection. These combinations can most easily be accommodated by a reverse-Hoogsteen geometry, shown below for the two most common pairings, A·U and A·A.

TABLE 2 Functional sequences present after three rounds of *in vitro* selection

Sequence		Number of isolates	Activity
5'	3'		
C A A A U	G U A A	6	0.9 (s.d. ± 0.3)
C A A A U	G G A A	2	1.0, 1.2
C U A A A	G A A U	1	0.8
C A A A A	G G A G	1	0.5
C A A A U	G A A C	1	0.5
C U A A U	G A A U	1	0.5
C A A A U	G A A G	1	0.5
C A A A U	G A G U	1	0.4
C U A A U	G A G A	1	0.3
U A A A U	A A A U	1	0.3
G U A A U	C A G C	1	0.3

RNAs from the third round of selection were cloned; 36 were sequenced and assayed as in Fig. 1, except that the assay was performed in the presence of 5 mM MgCl_2 and ribozymes (0.3 μM) were in excess over the labelled RNA oligomer. Under these reaction conditions the rate of ligation was linear for at least 15 min. Activities are normalized to the activity of the clonal wild-type ribozyme. Sequences having an activity $>25\%$ of the wild-type value are shown. Bases differing from the wild-type sequence are underlined; some sequences contain additional mutations outside the conserved core of the enzyme.

consistent with the existence of a reverse Hoogsteen base pair (Fig. 3). This hypothesis is also supported by site-directed mutagenesis experiments, and by the many A·A, U·C, and C·C pairings found in other group I introns, all of which can be most conveniently fitted to a reverse Hoogsteen geometry⁵. An alternative explanation of these data is that more than one geometry is permissible at this location⁶.

The existence of only A or U at J3/4-3 is consistent with mutational studies and with the hypothesis of a triple-helical scaffold proposed by Michel and Westhof⁷. This model proposes several structurally important base-triples, including an interaction between J3/4-3 and P6-1. Assuming the geometric equivalence of different nucleotide pairings, only an A or U can satisfy this triple interaction while P6-1 (which is outside the mutagenized region) remains G·U.

The system for *in vitro* genetic analysis that we have described is based on selection for ribozyme activity, and *in vitro* amplification. A similar system has recently been described by Robertson and Joyce⁸. These systems should have numerous applications. For example, they could be used to identify tertiary interactions, by selecting for second-site suppressors of deleterious mutations, and to select for mutations that confer altered substrate specificity. The methods can also be used, as we have here, to generate 'artificial phylogenies' covering limited regions of the enzyme, with the added advantage over natural phylogeny that all recovered sequences are known to be active. Finally, *in vitro* selection should be useful for exploring ribozyme activity under a wider range of conditions than is possible *in vivo*, and in selecting for novel catalytic activities. □

Received 7 June; accepted 16 July 1990.

1. Saiki, R. K. *et al.* *Science* **239**, 487-491 (1988).
2. Kwoh, D. Y. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 1173-1177 (1989).
3. Cech, T. R. & Bass, B. L. *A. Rev. Biochem.* **55**, 599-630 (1986).
4. Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. *Nature* **300**, 719-724 (1982).
5. Couture, S. *et al.* *J. molec. Biol.* (in the press).
6. Kim, S.-H. & Cech, T. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8788-8792 (1987).
7. Michel, F. & Westhof, E. *J. molec. Biol.* (in the press).
8. Robertson, D. L. & Joyce, G. F. *Nature* **344**, 467-468 (1990).
9. Burke, J. M. *et al.* *Nucleic Acids Res.* **15**, 7217-7221 (1987).
10. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. *Nucleic Acids Res.* **15**, 8783-8798 (1987).
11. Doudna, J. A., Gerber, A. S., Cherry, J. M. & Szostak, J. W. *Cold Spring Harb. Symp. quant. Biol.* **52**, 173-180 (1986).

ACKNOWLEDGEMENTS. We thank D. P. Bartel for many helpful discussions. This work was supported by Hoechst AG.